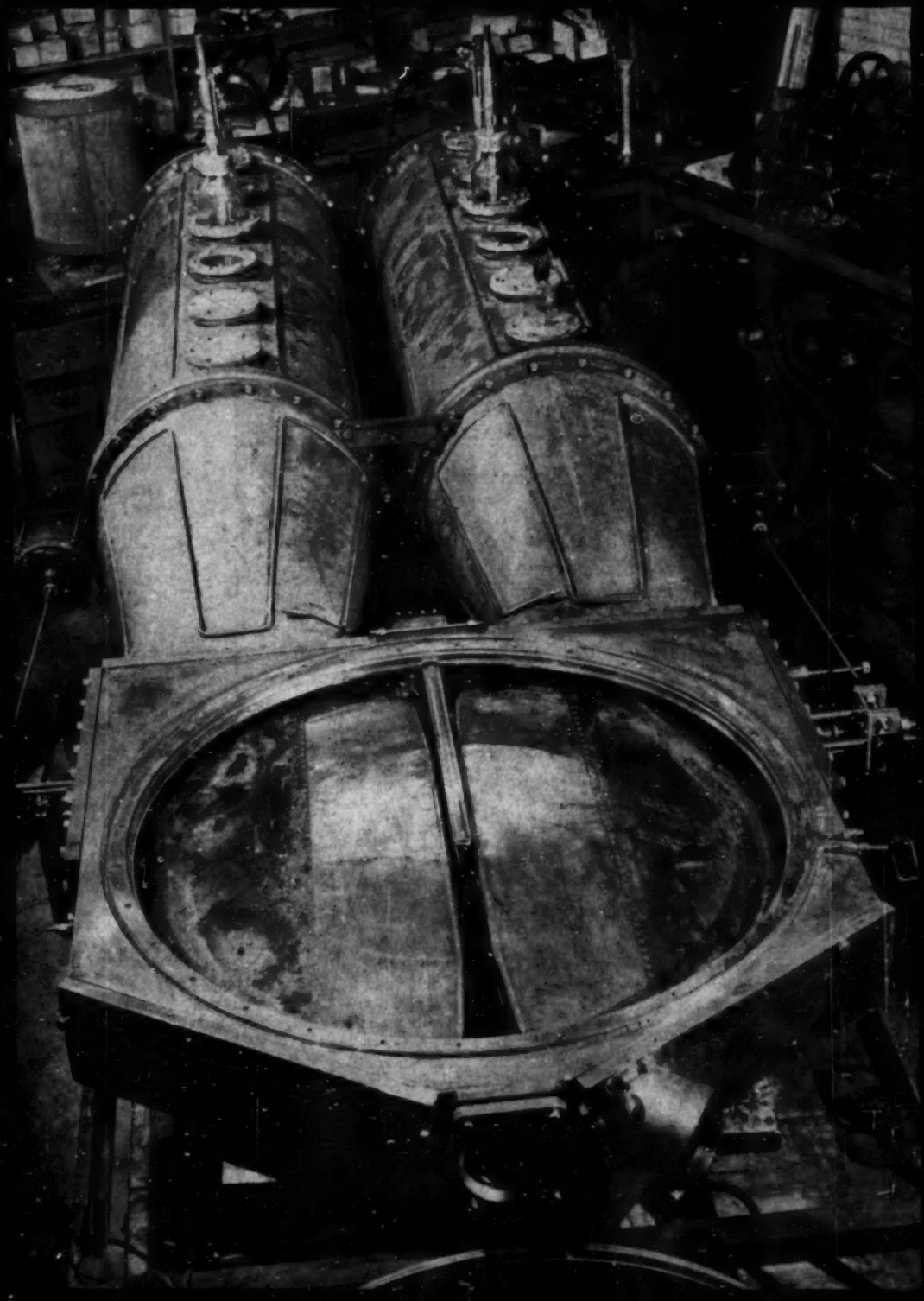
JOURNAL OF APPLIED PHYSICS



JOURNAL
OF
APPLIED PHYSICS

Vol. 15, No. 1 JANUARY, 1944

In This Issue

		Page
Volume Title Page		**I**
Applied Physics in the Postwar Period		**1**
The Cyclotron. I	M. STANLEY LIVINGSTON	**2**
Contributed Original Research		
Effect of Strain Rate Upon Plastic Flow of Steel	C. ZENER AND J. H. HOLLOMON	**22**
Problem of Reduction of Vibrations by Use of Materials of High Damping Capacity	ANDREW GEMANT	**33**
Technique of Microradiography	S. E. MADDIGAN	**43**
Antenna Theory and Experiment	S. A. SCHELKUNOFF	**54**
Modification of Hallén's Solution of the Antenna Problem	MARION C. GRAY	**61**
Exploring Stress-Strain Relations of Isotropic Plastic Solids	W. PRAGER	**65**
Studies in Three-Dimensional Photoelasticity. Stress Concentrations in Shafts with Transverse Circular Holes in Tension. Relation Between Two- and Three-Dimensional Factors	MAX M. FROCHT	**72**
In Every Issue		
Here and There		**20**
New Booklets		**20**
Recent Applications of Physics		**20**
Innovations in Instruments		**21**
Calendar of Meetings		**32**
Index to Advertisers		**xx**

BOARD OF EDITORS

ELMER HUTCHISSON,
Editor

Associate Editors

R. B. BARNES
OTTO BEECK
W. F. BUSSE
CLARK GOODMAN
SELIG HECHT
W. J. LYONS
MORRIS MUSKAT
THOMAS H. OSGOOD
PAUL E. SABINE
FREDERICK SEITZ
J. R. TOWNSEND

Publication of the American Institute of Physics

ADMINISTRATIVE STAFF

HENRY A. BARTON,
Director

MADELINE M. MITCHELL,
Publications Manager

JOHN T. TATE,
Adviser on Publications

GOVERNING BOARD

PAUL E. KLOPSTEG,
Chairman

GEORGE B. PEGRAM,
Secretary

T. D. COPE
R. B. DOW
L. A. DUBRIDGE
FLOYD A. FIRESTONE
HARVEY FLETCHER
R. C. GIBBS
K. S. GIBSON
ARTHUR C. HARDY
HERMANN MARK
A. G. WORTHING
JOHN T. TATE
N. W. TAYLOR
WALLACE WATERFALL

The JOURNAL OF APPLIED PHYSICS, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as PHYSICS. The JOURNAL OF APPLIED PHYSICS publishes editorials, news of physicists, and reviews, as well as technical papers of applied physics.

Manuscripts should be submitted to Elmer Hutchisson, Editor, 50th floor, 350 Fifth Avenue, New York 1, New York.

Subscriptions and orders for back numbers should be addressed to Prince and Lemon Streets, Lancaster, Pa. or to the American Institute of Physics, 175 Fifth Avenue, New York 10, N. Y.

Subscription Price

	U. S. AND CANADA	ELSEWHERE
To members of scientific societies	$5.00	$5.70
To all others	7.00	7.70

Back Number Prices

Complete set: Vol. 1, July, 1931–Vol. 13, 1942—**$80.50**
Yearly back number rate: **$7.70**
Single copies: **$0.70** each.

Changes of address and proofs should be addressed to the Publications Manager.

Advertising rates supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879.
Accepted for mailing at the special rate of postage provided for in the Act of February 28, 1925, authorized May 2, 1932.

Copyright 1944, by the American Institute of Physics.

Journal of Applied Physics

Volume 15, Number 1 **January, 1944**

Applied Physics in the Postwar Period

WITH the beginning of a new year, one's thoughts naturally turn toward the future and into the postwar period. We all realize that first of all the war must be pushed to a successful conclusion as vigorously as possible. We would be negligent, however, if we did not give some thought to the place which physics will be bound to take in postwar activities. That the postwar period is receiving much thought is evident from the number of plans now being brought forward. Sir William Beveridge, for example, has plans for postwar social security, the Pabst Breweries are offering $50,000 in awards for postwar employment plans, the Lincoln Electric Company is offering prizes to its employees for ideas on postwar planning, and undoubtedly many similar contests have been and will be initiated.

It is undoubtedly true that physics will play an important role in this period. A great awakening as to possible industrial applications of physics is already apparent. New divisions of the American Physical Society concerned with applied physics in electron optics, in rubber and textiles, and in aeronautics have been or are being organized. Large new laboratories are being established and many industrial organizations are advertising in newspapers as well as in technical journals for physicists. The thickness of this Journal, for instance, is increasing at a rate which is quite alarming in view of the WPB restrictions on paper. All of these things combined with the knowledge that many of the techniques used in implements of war will be just as useful in peacetime leave no doubt of the position which applied physics will occupy when peace comes.

There is another side of this problem, however, that needs some thought. Technological improvements are likely to contribute to unemployment unless at the same time there is an increased consumer demand for goods. Physicists should perhaps give some thought to the educational problem of creating a demand for a higher standard of living. If the effort now going into the war can be diverted into raising the general standard of living, there is not likely to be much of a serious manpower problem after the war is over. The fact that most of the developments that will enable the raising of the standard of living depend upon progress in physics puts some of the burden of making this higher standard desirable back into the lap of the physicists.

FOR FEBRUARY

Physics in 1943 ***Thomas H. Osgood***

Creep of Metals
S. Dushman, L. W. Dunbar, and H. Huthsteiner

Gaseous Diffusion in Metals
H. S. Coleman and H. L. Yeagley

The Cyclotron. II
M. Stanley Livingston

Contributed Original Research papers on electroplating, antennas, velometry, and electron microscopy.

The Cyclotron. I*

BY M. STANLEY LIVINGSTON

Massachusetts Institute of Technology, Cambridge, Massachusetts

The cyclotron is in process of a metamorphosis from an academic luxury for nuclear physicists to an indispensable tool for applied science and industry. Tracer experiments with induced radioactivities have become invaluable for basic studies in other sciences and as a guide to engineering development in many industries. The radioactivities and neutron radiations themselves are being put to use in industry and in medicine on an ever increasing scale. The time is near when the use of the cyclotron as a production unit may even outweigh its further scientific usefulness. Cyclotron physicists, well aware of this trend, are eager to push the development of the instrument to engineering perfection. To further this effort, this paper evaluates available information which may aid in improving the efficiency, reliability, and output and in determining most economical sizes and designs of cyclotrons for large scale applications.

TABLE OF CONTENTS

1. Introduction
2. Physical Principles
3. Electric and Magnetic Focusing
4. Maximum Energy
5. Oscillator Circuit
6. Magnet
7. Resonant Circuit and Chamber
8. Ion Source
9. Shimming
10. Deflector
11. Target Arrangements
12. Vacuum Pumps
13. Vacuum Seals and Techniques
14. Controls
15. Laboratory Design
16. Shielding
17. Cyclotron Costs
18. Yields of Nuclear Reactions
19. Optimum Size
20. Conclusions and Results

1. INTRODUCTION

THE history of the early development of the cyclotron can be followed through the publications and progress reports of the University of California Physics Department.[1–7] From small beginnings with modest apparatus the group expanded rapidly and continuously changed with the continuing success of the development, each generation adding to the techniques and improving and enlarging the instruments. The inspirational direction of Professor Ernest O. Lawrence was a major factor in accomplishing this rapid and successful development, which has culminated in the present Radiation Laboratory of the University of California. Meanwhile in many other laboratories cyclotrons were constructed, at first largely designed by graduates of the Berkeley school. Soon these laboratories were able to make important contributions to the development. A valuable factor in this phase was the free interchange of information between laboratories, by private visits and correspondence, exchange of blueprints, and by mimeographed progress reports sponsored by the Berkeley group. This healthy exchange of ideas and designs has made the modern cyclotron a composite product of many laboratories and scores of individual contributions. A large majority of the technical developments on the cyclotron have not been described in published reports, but after alterations and improvements as they passed from one laboratory to another, have become part of the general store of technical information.

In the discussion of designs and techniques to follow, in which the M.I.T. cyclotron is described, full recognition of the indebtedness to other laboratories and the accumulated store of technical knowledge is intended. Wherever possible, references are given to the individuals or laboratories originating an idea or development. During the several years of construction and development at this laboratory the cooperative efforts of many individuals were involved, and all have contributed to the progress. Both in design and in subsequent development serious attempts have been made to keep step with

* This is the first of two parts of a comprehensive paper on the cyclotron. The second part including Sections 11 to 20 will appear in the February issue of this Journal.

improvements in other laboratories and to utilize the best ideas and techniques available. In addition, we have attempted to reduce the unreliability inherent in most laboratory-built equipment by a maximum use of commercially available apparatus and materials. Although there may be valid criticisms of some of our conclusions, we feel that most of the features described are representative of present best practice and maximum efficiency.

2. PHYSICAL PRINCIPLES

The cyclotron is an instrument for the production of high velocity particles to be used as projectiles for nuclear disintegration, without the use of high voltages. Ions produced in a low pressure gaseous discharge are accelerated by radiofrequency electric fields between two semicircular hollow electrodes called "dees" because of their shape. The dees are symmetrically located within a vacuum chamber between the poles of an electromagnet. After traversing a half-circle within one of the hollow electrodes, the ions return to the diametral gap between electrodes. For the condition of resonance, the magnetic field is adjusted so that the time required for an ion to complete a half-circle is equal to the time for reversal of the radiofrequency electric field. So the ion experiences another acceleration, acquires a higher velocity, and traverses a path of larger radius within the other electrode. The angular velocity is a constant in the uniform magnetic field, the time required for a semicircular path being independent of ion velocity. So the ions are accelerated each time they cross the gap, traveling in ever-widening circles until they reach the periphery of the electrodes. See Fig. 1. The final ion energy is the sum of the individual accelerations, or can be computed from the radius of the last semicircular path in the magnetic field.

The force on an ion of mass m, charge e, and velocity v moving at right angles to a magnetic field H is perpendicular to its direction and produces motion in a circular path of radius r with a radial acceleration v^2/r:

$$Hev = mv^2/r \quad \text{or} \quad v = (e/m)Hr. \tag{1}$$

For the condition of resonance, the angular velocity must be equal to $2\pi f$ where f is the frequency of the applied electric field:

$$v/r = 2\pi f \quad \text{or} \quad H = 2\pi(m/e)f = 1.312f, \tag{2}$$

where H is in kilogauss and f in megacycles for deuterium ions (deuterons). This linear relation determines the magnetic field for resonance at any frequency, and is plotted in Fig. 2.

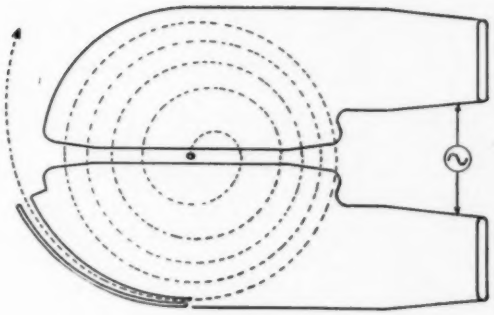


FIG. 1. Schematic diagram of the cyclotron showing the multiple accelerations of an ion within the electrodes and the electrostatic deflection of the emergent beam.

The energy of an ion of velocity v, in terms of the equivalent accelerating potential E, is:

$$\tfrac{1}{2}mv^2 = Ee \quad \text{or} \quad E = \tfrac{1}{2}(m/e)v^2. \tag{3}$$

Combining (1), (2), and (3) we obtain the equivalent energy for an ion in a path of radius R:

$$E = \tfrac{1}{2}(e/m)H^2R^2 = 2\pi^2(m/e)R^2f^2 = 0.0965f^2 \quad (\text{for } R = 19.0 \text{ inches}). \tag{4}$$

This relation is also plotted in Fig. 2 in Mev units for the 19-inch maximum radius of the M.I.T. cyclotron. With an accelerating voltage $2V$ between electrodes and for n accelerations:

$$E = 2nV \quad \text{or} \quad r = 2\left(\frac{m}{e}\frac{V}{H}\right)^{\frac{1}{2}}\sqrt{n}, \tag{5}$$

so the radii of successive paths vary with the square roots of integral numbers. With 140 kilovolts between electrodes for 15-Mev resonance conditions, for example, the radius of the first semicircular path (accelerated through 70 kilovolts) would be the order of 1.3 inches and the difference in radii of the last two paths would be the order of 0.18 inch.

For protons, which have a charge to mass ratio twice that for deuterons, the magnetic field for

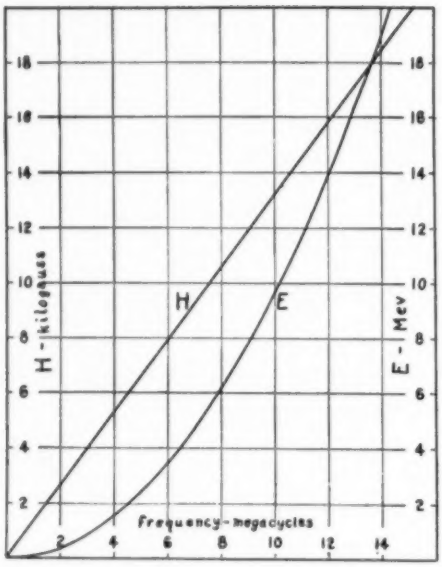


FIG. 2. H—magnetic field *vs.* frequency of electric field for cyclotron resonance with deuterium ions. E—deuterium ion energy *vs.* frequency for 19.0-inch maximum radius of path in the magnetic field.

resonance at a given frequency will be half the value for deuterium ions plotted in Fig. 2; the energy will also be half that for deuterons at the same frequency. For equivalent magnetic fields the applied frequency is doubled (often unattainable for practical reasons) and the final proton energy will be twice that for deuterons. Singly-charged hydrogen molecular ions and doubly-charged helium ions have the same e/m as deuterons (differing by only one part in 1300 and 1600, respectively). The resolution of the cyclotron as a mass spectrometer is just sufficient to separate the three resonance peaks at the same frequency. Hydrogen molecular ions have the same total energy, but should be considered as protons of half the deuteron energy in nuclear reactions. Due to their double charge, helium ions have twice the deuteron energy for similar resonance conditions.

3. ELECTRIC AND MAGNETIC FOCUSING

The feature which makes the method of multiple accelerations practical is the focusing resulting from the electric and magnetic fields. In crossing the gap between electrodes in an accelerating electric field, the ions experience a deflection toward the central plane, a one-dimensional focusing by the "cylindrical" electric lens.[4] The magnitude of the electric focusing can be described by the relation:[8]

$$\Delta\alpha = -\left(\frac{eV_0}{E}\right)\left(\frac{\omega}{v}z\right)\sin\theta \cdot -\tfrac{1}{2}\left(\frac{eV_0}{E}\right)^2\left(\frac{\lambda}{k}z\right)\cos^2\theta, \quad (6)$$

where $\Delta\alpha$ is the deflection at a distance z from the central plane of an ion of charge e, average energy E, and corresponding velocity v, crossing a sinusoidal electric field of frequency $\omega/2\pi$ at a phase θ relative to the maximum potential $2V_0$. The first term describes the time variation of electric field as the ion crosses the gap. The deflection is toward the central plane on approaching the gap and away on leaving. If the field were constant these would exactly balance, but if the field is decreasing (positive phase θ) the result is a focusing. The second term introduces the variation of particle energy in crossing the lens. The increased velocity of the accelerated ion when crossing the defocusing part of the field allows the forces to act for a shorter time, also giving a focusing. The λ and k terms are geometrical constants determined by the physical dimensions of the gap. The net result is a displacement of path toward the central plane and a reduction of the angular divergence, indicated schematically in Fig. 3. It can be expressed in terms of the focal length of the lens which is reasonably short for low energy ions in their first few accelerations, where V_0 is appreciable compared with E, but becomes very long for higher energy particles. Each ion traverses the lens the order of 100 times. The deflection for all

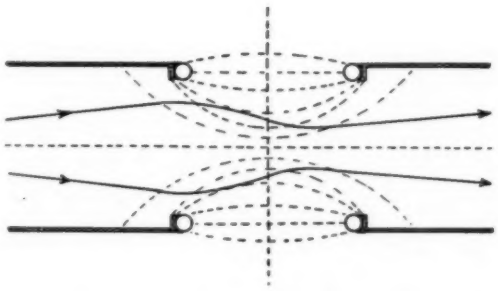


FIG. 3. Cross section of dees near the acceleration region showing schematically the mechanism of electric focusing.

but the first few accelerations will be primarily due to the time variation of the field which is independent of the geometry of the gap. If we visualize the spiral path of the ions in the magnetic field as unrolled into a straight line, the successive lenses, of increasing focal length, would be spaced at increasing intervals proportional to the square roots of integers. The paths of ions through such a combination of lenses can be expressed as a function of n, the number of accelerations.[8]

$$z=c\left(\frac{E}{\sin\theta}\right)^{\frac{1}{4}}\sin\left[\pi^{\frac{1}{2}}\int_1^n\left(\frac{eV_0}{E}\sin\theta\right)^{\frac{1}{2}}dn+\delta\right], \quad (7)$$

where c and δ are integration constants. If the phase θ is positive (field decreasing as ions cross the gap), this represents an oscillation of ions about the median plane with long wave-length having an envelope diverging slowly with $E^{\frac{1}{4}}$. If the phase is negative (increasing field), the motion is exponential and the ions are strongly defocused. So for positive phase, the electric fields produce fairly strong focusing for the first few accelerations, and weaker focusing with a slowly diverging envelope for the rest of the ion paths. In the absence of magnetic focusing, the envelope would intersect the top and bottom surfaces of the electrodes and only a small fraction of the resonant ions would be near the central plane at the exit slit.

In an absolutely homogeneous magnetic field, in which field lines were strictly vertical and parallel, there would be no vertical deflections and no focusing. The focusing results from a slight deviation from homogeneity consisting of a field decreasing radially outward so that lines of force are concave toward the center.[4] This is the natural shape of magnetic field near the periphery between cylindrical pole pieces and is called fringing. An adequately decreasing field can be obtained by shims (see Section 9) in the central region. The total deviation should be small, of the order of 1 percent, in order to maintain resonance, and field lines should be symmetrical about the central plane. The small radial component of magnetic field off the central plane acts to deflect ions back toward this plane. This is indicated schematically in Fig. 4. For

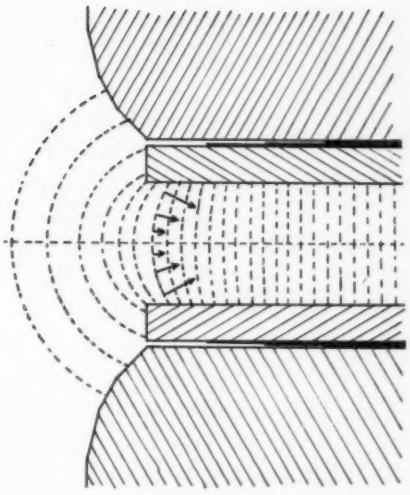


FIG. 4. Representation of a fringing magnetic field and the focusing forces acting on ions off the central plane due to lines of force concave toward the center of the chamber.

such a focusing field the equation of motion of an ion in suitable cylindrical coordinates shows an oscillation about the median plane similar to that for electric focusing. The amplitude of the oscillations, z, is given by:[8]

$$z=-c\left(\pi^2\frac{d\log H_z}{d\log r}\right)^{-\frac{1}{2}}, \quad (8)$$

where H_z is the field component producing motion in the z direction at radius r. The amplitude will decrease continuously wherever a focusing field exists, at a rate determined by the magnitude of the radial component of the field. If the field is convex inward, the ions will be strongly defocused.

The actual motion of the ions will be due to the combined action of electric and magnetic fields, and adequate focusing will exist if the maximum amplitude of the envelope enclosing ion oscillations is less than half the internal height of the electrodes. Experimental observations[9] of the distribution of ions with respect to distance from the median plane at different radii are in agreement with the theoretical calculations. They show a broad spread at small radii which becomes narrower or more focused as the radius increases. The electric focusing is predominant during early accelerations where the ion energy is low; the magnetic focusing dominates for large

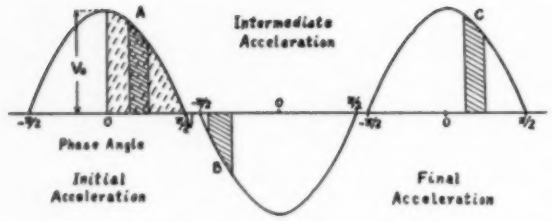


FIG. 5. Time variation of high frequency dee voltage indicating the allowed phase shift for ion accelerations from initial to final accelerations.

radius and high velocity ions. They are of the same order of magnitude at about $\frac{1}{3}$ the final radius and it is in this region where the envelope of ion paths has its largest amplitude. The dees can be made to conform roughly to this shape, thereby minimizing electrical capacitance. The effective height of the source is limited by the relative weakness of the electric focusing, so a narrow source on the median plane (the arc source described later) is more effective than a source the full width of the electrode aperture for equivalent ion outputs. Furthermore, only those ions from the source making small angles with the median plane can be focused. This limitation is partially overcome by extensions on the electrode edges near the source (see Fig. 11) to produce a stronger focusing in the first acceleration.

The radially decreasing magnetic field required for focusing causes a shift in the phase at which ions cross the acceleration regions relative to the radiofrequency cycle. The magnetic field at the center of the chamber will be slightly larger than a theoretical uniform "resonance field," while the field at large radii will be smaller. This will cause the ions first to lead the voltage phase and then to lag behind it. During the 100 or more accelerations the resultant migration in phase may be over an appreciable portion of the radiofrequency cycle. They must start during a positive phase angle θ in an accelerating field, limited to the light shaded area of the radiofrequency cycle shown in Fig. 5. At least during the early accelerations where electric focusing predominates they must remain in this region, but can have a negative phase after the magnetic focusing is large enough to overcome any defocusing by the electric field. In practice iron shims are used to produce some magnetic focusing at quite small radii so the ions can start at a small positive phase where the high electric field gives maximum emission from the ion source. This is indicated by the band labeled A in Fig. 5. As long as the magnetic field is greater than the resonance field they will gain in phase, but must not exceed the limit of $-\pi/2$ beyond which they would be decelerated. This is shown on the intermediate acceleration cycle of Fig. 5 as the band B. The ions will start to lag in phase when the magnetic field is less than that required for resonance and will return to phase $\theta=0$. They should have large final accelerations (phase near zero) so that the difference in radius of successive circular paths ΔR will be large enough for them to pass behind the usual septum at the exit slit and enter the deflector channel. Only such ions can be deflected into an external beam (band labeled C in Fig. 5). So the practical maximum migration in phase will be from a small positive value to $-\pi/2$ and back to a small positive value. The internal resonant ions observed on a probe target (see Section 11) do not have this ΔR limitation and can have the phase lag extending to $+\pi/2$, so larger beam intensities can be anticipated on such a target.

4. MAXIMUM ENERGY

The relativistic increase in mass of the ions at high energies would need a magnetic field increasing radially outward to maintain resonance. This is forbidden by focusing requirements, so the ions will fall behind in phase more rapidly at large radii. The necessity for keeping the total phase change within the allowed range puts a severe limitation on the radial decrease in magnetic field desired for focusing. A useful rule-of-thumb is that the fractional decrease in field $\Delta H/H$ should not exceed the reciprocal of the number of ion accelerations $1/n$ (i.e., 1 percent decrease for 100 accelerations). The phase shift can be reduced by increasing accelerating voltage $2V_0$. This widens the band which will be within the phase limits, and in general results in higher ion beam intensities. The maximum energy obtainable with a cyclotron under these conditions can be computed from the magnetic field which provides the minimum necessary focusing and the maximum allowable phase shift.

The maximum energy is given by the relation:[8]

$$E_n = 1.54(2V_0AZ \sin \theta_0)^{\frac{1}{2}}, \quad (9)$$

where A is the atomic weight, Z the atomic number of the ions, θ_0 is the initial phase, and E_n is expressed in Mev and V_0 in kilovolts. *This maximum is proportional to $V_0^{\frac{1}{2}}$ and so is determined only by the practical limit of electrode voltage.* For deuterons this becomes:

$$E_n = 3.08V_0^{\frac{1}{2}}(\sin \theta_0)^{\frac{1}{2}}. \quad (9a)$$

Experience in this laboratory shows that peak potentials between dees of 80 kilovolts for 12-Mev ions and 140 kilovolts for 15-Mev ions are adequate to give reasonable beam intensities. For maximum beam currents even higher voltages are needed. These observations require an initial phase angle of 22° and 20°, respectively, justifying the statement made previously that small positive values would be needed to obtain a satisfactory ion source emission. Using the observed value at 15 Mev to determine a phase constant for the M.I.T. cyclotron, the equation above becomes:

$$E_n = 1.8V_0^{\frac{1}{2}}. \quad (9b)$$

This indicates that potentials of 200 kilovolts between dees would be required for 18-Mev deuterons, for example.

The high frequency power requirements are closely proportional to the square of the electrode potential in a particular circuit, so will increase with the fourth power of the ion energy setting. For the circuit described in this paper the input oscillator power to obtain the 70 kilovolts to ground on each dee required for 15-Mev ion beams is 50 kilowatts. (Estimating the over-all oscillator efficiency to be 50 percent this provides 25 kilowatts for the dee circuits and resonant ion load.) So for our circuit: P (kilowatts) $= 0.001E^4$, suggesting a power requirement of 105 kilowatts at 18 Mev. This is within the maximum ratings of the oscillator and power supply and indicates the possibility of operation at that energy.

To reduce power requirements and increase the maximum energy the dee circuits should have larger conductors and lower resistance. Resistance would vary inversely with the diameter of the coaxial dee lines, to a first approximation. Still higher energies could be obtained by increasing the magnet diameter and all other dimensions proportionately. Since ion energy varies directly with the square of pole diameter, the power for equivalent dee potentials in cyclotrons of different size will be inversely proportional to the square root of the energy. The over-all power requirements will therefore increase with $E^{7/2}$. This is an extremely rapid variation and implies that power requirements will be the limiting factor governing the ultimate maximum energy obtainable from the cyclotron.

Several suggestions have been made for maintaining focusing even in the radially increasing magnetic fields needed to preserve resonance with the relativistic increase in mass. This would result in much lower dee potentials for the allowed phase variation and so decrease power requirements. One which has been worked out in detail uses a variation of magnetic field with polar angle.[10] Other suggestions include the use of quadrants rather than dees, shaping of the opposing faces of the dees or quadrants to follow the phase shift, and use of electrical grids within the electrodes to obtain auxiliary electric focusing. The application of one or more of these techniques may eventually result in raising the maximum energy limits without the excessive power indicated by the foregoing analysis.

5. OSCILLATOR CIRCUIT

The supplying of adequate radiofrequency power to the resonant circuit of the cyclotron is a problem in radio engineering. Techniques and methods developed in the radio industry for broadcast stations are not as readily applicable as could be hoped, due to the unusual characteristics of the cyclotron as a load. The problem is to maintain a high frequency (10 to 20 megacycles) potential difference of 100,000 to 200,000 volts between the opposing faces of the dee-shaped electrodes. To transform from the potentials available with conventional high frequency generators requires the use of resonant circuits. Considerations of electrical efficiency and mechanical rigidity have led to the use of self-supporting coaxial quarter-wave resonant lines of large diameter in vacuum. This design eliminates the mechanical insulating supports required

in earlier cyclotrons. It is shielded from radiative losses and has the high Q requisite to obtaining the necessary high potentials for a reasonable power input. A satisfactory alternate would be a shielded pair arrangement. Each dee has an electrical capacitance to ground of about 2×10^{-10} farad. To produce a potential to ground of 70,000 volts at 12 megacycles involves a current of more than 1000 amperes. In a circuit having a high frequency resistance that is estimated to be 0.01 ohm this represents about 10 kw of power in each of the two dee lines. An additional requirement of the order of 10 kw is involved in acceleration of the circulating ions and the resistive losses due to the increased current. This power may be supplied at any desired impedance level by entering the resonant dee lines at the appropriate distance from the nodal end or by inductive or capacitative coupling of suitable magnitude.

The high frequency generator to supply this power must meet certain requirements. The power must be supplied at the resonant frequency of the dee lines. The phases of the two dee lines must be in opposition (push-pull) to give maximum potential between dees. Sparks and heavy gaseous discharges are common, particularly during tune-up or starting periods. They are essential to the proper electrical conditioning of the electrodes and vacuum chamber, and sufficient power must be fed into them to outgas the surfaces. In discharges the effective impedance of the resonant circuit is reduced to a very low value. Protective features should be incorporated in the design to prevent the oscillator from destroying itself under the load changes resulting from discharges. Thermal stresses set up in the electrodes by the high frequency heating result in some physical distortion of the electrodes, even for the best of designs. In such a high Q circuit, the slight frequency changes resulting from the variable load of resonant ions and discharges are significant. The generator must follow such frequency changes to maintain maximum electrode potentials.

A self-excited oscillator tighly coupled to the load will follow small frequency changes, and is used as the basis of most cyclotron high frequency generators. Two power oscillator tubes are usually arranged in push-pull to take advantage of the symmetry and simplicity of the circuit and to double power output. Such an oscillator normally includes resonant grid and plate circuits. Power is transmitted to the load from the plate (tank) circuit by means of transmission lines of appreciable length. In order to obtain efficient power transfer the circuits are coupled at optimum impedance, resulting in multiple resonances. Any pair of coupled circuits individually tuned to the same resonant frequency exhibits two modes of oscillation depending on whether they are in phase or out of phase at the mutual coupling element. The out-of-phase or push-pull mode is usually the desirable one, as in the case of the two dee circuits. In the self-excited oscillator there may be resonant circuits in the grid, the plate, and the transmission lines as well as the cyclotron circuits, all closely coupled and tuned to the same frequency. The generator must include circuits or tuning elements to suppress the undesirable frequencies. One such problem of major importance is that of transfer of power from the tank circuit to the high Q dee circuit. If we simplify the problem by neglecting the coupling units and transmission lines, we can consider the tank circuit as directly coupled to the load circuit. The undesirable push-push mode of oscillation in this case results in no power transfer to the load. A self-excited oscillator will oscillate in the mode which stores the maximum energy in the associated grid and plate circuits for a given energy output. If this energy storage is insufficient (low Q for grid and plate circuits), the oscillator will tend to unload itself as peak tuning is approached, a condition known as snap-tuning.

The problem has been approached differently in the several laboratories, conditioned by engineering advice and the paralleling development of the cyclotron resonant circuit. In early models, when the dees were supported on stems through glass insulators and the inductance consisted of a coil of copper tubing across the terminals, relatively simple self-excited oscillators were adequate. A two-tube, push-pull, tuned-plate, tuned-grid circuit became standard. It was mounted close to the cyclotron and direct coupled so that no transmission lines were

needed, the cyclotron resonant circuit was also the tank circuit, and the grid coil was tuned to resonance. The development by the Franklin Institute group[11] of a parallel-line inductance to replace the coil resulted in much higher Q values. Three-inch copper pipes spaced several diameters apart and with an adjustable shorting bar or node were added as extensions of the dee stems. A further improvement by the Columbia University group introduced quarter-wave, coaxial, evacuated dee lines. With these extensions and the increasing dimensions of the cyclotron, it became necessary to use larger and more powerful high frequency generators and transmission line coupling to the load. Several basically different designs have resulted which will be described and criticized separately.

Master Oscillator-Power Amplifier:[12] The communications industry has developed generators for the frequency and power ranges required by the cyclotron, usually for driving a fixed load antenna. Such units consist of a fixed frequency master oscillator and a final amplifying stage operating at the desired power level driven by the oscillator or by suitable intermediate amplifiers. As adapted to cyclotron practice, the master oscillator is frequency controlled by a feedback from the cyclotron resonant circuit so that it follows small variations adequately. Discharges tend to overload the power amplifier tubes. However, most of the problems of protection have already been solved in communications practice. With further development for cyclotron applications, this circuit, though relatively complicated, may well prove to be worth the effort.

High Q Grid Circuit:[13] The conventional self-excited oscillator has been made to render satisfactory service by using a highly selective grid circuit to limit the range of frequencies for which the grid circuit will provide the proper exciting voltage. The high Q is desirable both to increase the stored energy and to give a capacitative impedance at the frequency of the undesirable mode of oscillation where no power is transferred to the load. An operating Q of 400 to 500 is obtained by large diameter (6 inch) parallel conductors in the push-pull grid circuit. This is still small enough to cover the normal range of frequency variation of the dee circuit. Such highly resonant grid lines increase the probability of high frequency parasites, making necessary additional complications for their suppression in the form of wave traps. In order to couple the tank circuit to the dees closely enough to follow frequency variations, half-wave transmission lines are used. Such lines are equivalent to a direct connection in that they show at one end whatever impedance is applied at the other end. Accordingly, they enter the dee lines at the relatively high impedance level of the oscillator anodes (several thousand ohms). The high impedance tends to increase the frequency difference between the two modes of oscillation, making it easier for the grid circuit to discriminate between them. The half-wave line provides excellent coupling for power transfer, but suffers from the fault that the low impedance of the resonant circuit in discharge conditions is imposed directly on the tubes and tends to overload them. A constant current network in the power supply is used to limit the power available for such discharges and so protect the tubes. For three-phase, high voltage power supplies, this is an expensive form of protection, although it is essential to this circuit. There is a question as to the degree of protection afforded, since the time of operation of the control network (the order of the power frequency cycle of 1/60 sec.) includes thousands of cycles at the oscillator frequency, and may permit dangerous peak currents to flow which might shorten tube life.

One of the most valuable developments of the Berkeley laboratory has been the demountable power oscillator tubes. They are constructed in the machine shop with water-cooled copper anodes, grids, and filament stems, and are continuously pumped with separate oil diffusion pumps. After years of development to higher power and more reliable vacuum techniques, they are now quite satisfactory. Reliable operation for periods of up to a year between filament replacements and efficiencies equivalent to commercial sealed-off tubes can be expected. The high initial cost of construction, involving a well-equipped machine shop and experienced supervision, may influence a decision to use commercial sealed-off tubes, as at M.I.T. Over years of operation the saving in replacement

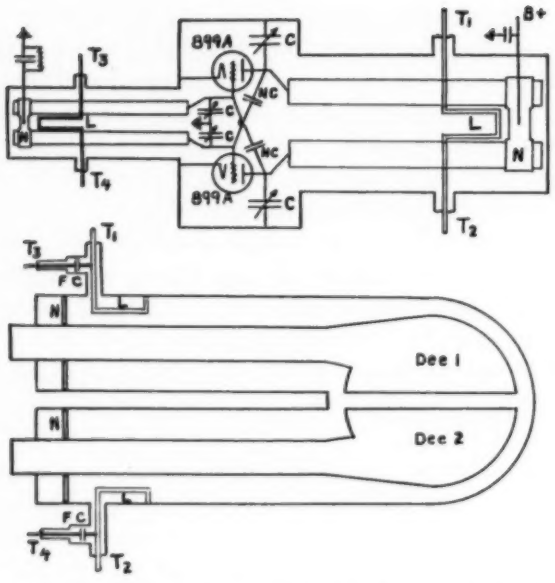


FIG. 6. Schematic representation and circuit diagram of the neutralized oscillator and dee circuits. Not to scale. *T*—transmission lines, *N*—adjustable nodes, *L*—coupling loops, *C*—tuning capacitors, *NC*—neutralizing capacitors, *FC*—feed-back capacitors. The plate and grid circuits are shielded pair lines and the dee circuits coaxial lines.

costs should make the demountable tubes more economical.

Neutralized Oscillator:[14] The circuit developed for the M.I.T. cyclotron is a neutralized oscillator with feed-back excitation from the load. The two RCA 899A power tubes are arranged in a push-pull, tuned-plate, tuned-grid circuit, shown diagramatically in Fig. 6. The plate circuit is a shielded pair line of 3-inch conductors which also carries the water cooling for the anodes. It has a sliding node and variable condensers to vary both inductance and capacitance, so the L/C ratio can be adjusted as well as frequency. This adjustment can be very useful in elimination of high frequency parasitic oscillations. The grid circuit is similar but of $1\frac{1}{2}$-inch conductors and is self-biased through a 5000-ohm resistor. Transmission lines of 72 ohms characteristic impedance, of non-critical length due to terminating each end in an inductive coupling loop of the same impedance, are used to transfer power from the plate circuit to the dee circuits. The physical length of the lines is of the order of $\frac{3}{16}$ wavelength. Without neutralization the generator operated as a conventional self-excited oscillator due to feedback through inter-electrode capacitances, and the relatively low Q grid and plate circuits allowed snap-tuning into the undesirable no-load mode of oscillation. To avoid this difficulty, the tubes are carefully neutralized with vacuum condensers between opposite grid and anode terminals, and excitation power is fed back from the load circuit to the grid circuit through smaller 72-ohm transmission lines, also terminated in 72-ohm loop couplings. A point was sought in the chain of circuits at which a voltage maximum for the desirable mode of oscillation would occur and at which little or no voltage of the undesirable modes would be present. The most satisfactory point seems to be at the end of the coupling loop in the dee lines where the plate transmission lines terminate. Appreciable voltages occur here only for the two resonant frequencies of the double-dee system. These two modes differ by about 500 kilocycles and the grid and plate tuning is sufficiently sharp to tune out the push-push mode. This point also has the advantage that it shows an appreciable voltage of the correct frequency even though discharges within the chamber have reduced dee voltages to near zero, so that excitation is maintained and power is fed into the discharges. An alternative point for connecting the grid feed-back transmission line would be at a separate dee line coupling loop; here, however, the excitation voltage is a strong function of dee voltage and drops sharply on discharge, thus reducing power to discharges beyond the desirable point. A radiofrequency power limiting feature is inherent in this feed-back system. Discharges drop the impedance of the dee circuit to near zero and reduce impedance of the dee line coupling loop to about 50 percent, reducing the excitation voltage transmitted to the grids. This results in a decrease in high frequency power in the plate circuit and a corresponding reduction in d. c. power input to the tubes. So the necessity of a constant current network in the power supply is eliminated.

The circuit behaves much like a Class C power amplifier stage in tuning manipulations. The plate and grid circuits are tuned to give the minimum plate current and the maximum grid excitation. Feed-back condensers and coupling loops, which control the r-f excitation, are ad-

justed to make the d. c. grid current and voltage match the rated impedance of the tubes. The oscillator has a higher efficiency and better stability with a slight excess of grid excitation. A half-wave rectifier in the grid return circuit prevents the grids from going positive during tuning manipulations. Thermal measurements of power in the dee circuits have indicated an efficiency well in excess of 50 percent at 40 kilowatts input. Slight detuning of the grid circuit will result in large phase variations and is used to balance out the retardation in phase introduced by the transmission lines and coupling loops. Improper adjustment of the coupling loops can give resonances in the transmission lines and may cause damage at voltage or current maxima, but can be controlled if lines are constructed to be well away from the critical quarter- or half-wave-lengths. The most objectionable feature of the circuit lies in the multiplicity of adjustments and tuned circuits. With so many possible variables, the manipulations can be time consuming. The answer is in careful calculation or measurement of optimum settings for all but the basic grid and anode tuning variables. The circuit will also operate satisfactorily with only one plate and one grid transmission line, leading to the same dee line loop. This reduces the number of possible resonant circuits.

In summary, the neutralized oscillator circuit is found to have the following properties:

(1) Power is supplied only at the resonant frequency of the dee system.

(2) All modes of oscillation except the two fundamental modes of the double-dee system are blocked by shielding and cross neutralization of the push-pull oscillator. No oscillations of any type are observed if the excitation is disconnected and no parasitic frequencies have been observed to date.

(3) Proper push-pull operation of the dee circuits is favored by symmetrical loops with center taps by-passed to ground coupling the push-pull oscillator and the load. The two modes can be readily separated by the plate and grid tuning variables.

(4) The r-f power limiting feature of the feedback system protects the tubes against damage during discharges in the chamber.

(5) Positive power feed to the load is maintained through showers of sparks and dense blue glow discharges.

(6) The oscillator tubes have a sufficiently high amplification factor (40 to 80) to be protected against loss of excitation.

(7) Inductive loop coupling to the plate and grid circuits allows transmission lines and dee circuits to be at d. c. ground potential.

(8) Transmission lines terminated in their characteristic impedance have a uniform voltage and current distribution, can be of small diameter, and are of uncritical length.

(9) Low impedance loops and lines reduce the problem of voltage insulation in the transmission lines and at the point of entry to the vacuum system.

6. MAGNET

The magnetic circuit for the M.I.T. cyclotron was designed by Mr. G. H. Cole of the American Rolling Mills Company.[15] The guiding principle was the use of a minimum of high quality magnetic iron consistent with a simple structural design. Armco Ingot Iron, with its excellent magnetic properties, was used in the form of large castings, admirably free from blow holes and having magnetic properties essentially identical with those of the annealed wrought iron. A maximum field of 18,000 gauss was desired in a $5\frac{1}{2}$-inch total gap between pole faces of 42 inches diameter. Tapering of the poles and the cross-sectional areas of the magnetic circuit were determined from the computed fringing fields and a flux density limit set at 20,000 gauss within the iron. As shown in Fig. 7, the magnet is formed of six castings consisting of two cylindrical, tapered poles, two yokes, and two rectangular uprights. The yokes were slightly tapered and cast with bevelled external corners and conical recesses opposite the pole locations to conserve iron. Surfaces of contact and pole faces were machined; in the final assembly the pole faces were flat and parallel to 0.0015 inch. Dowel pins were used for alignment and one 3-inch bolt was used to support the upper pole. The magnetic circuit weighs approximately 75 tons.

The exciting coils were designed and constructed to specifications by the General Electric

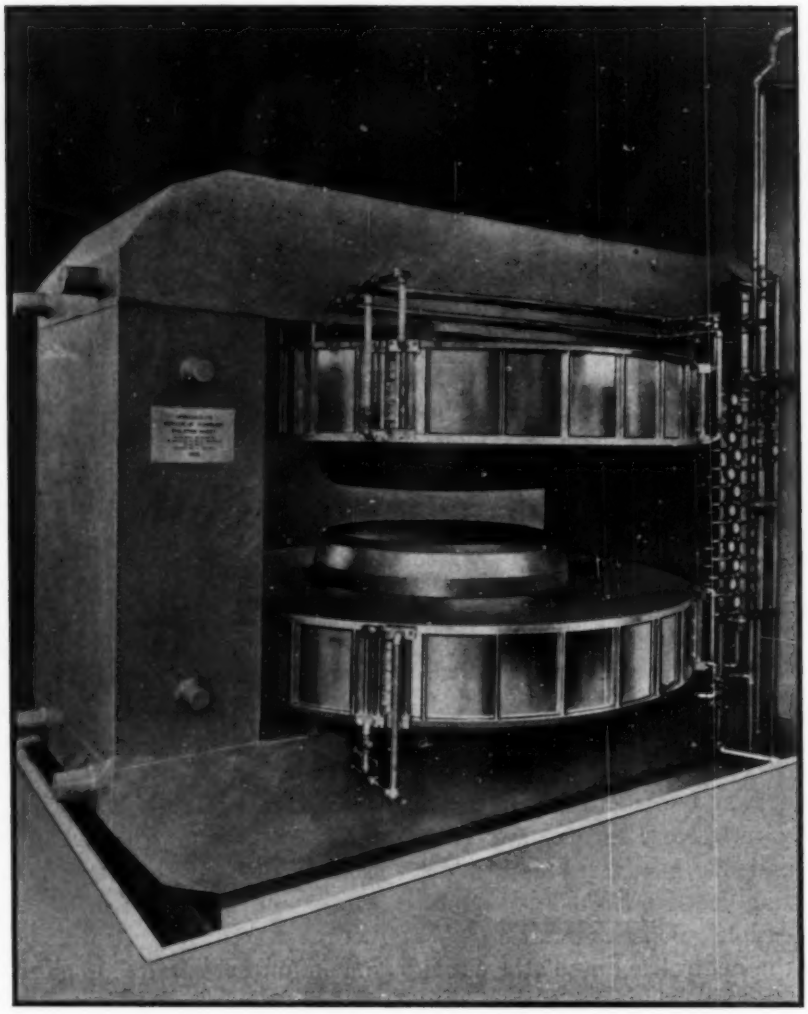


FIG. 7. Massachusetts Institute of Technology cyclotron magnet assembly showing the cylindrical, tapered poles, the magnetic circuit, and exciting coils.

Company. Copper strip $1\frac{1}{4} \times \frac{1}{8}$ inch cross section with slightly rounded edges was wound in flat layers with a double layer of thin insulating paper between turns. There are eight layers totaling 976 turns in each coil. Between alternate layers of electrical conductor a layer of rectangular copper tubing is provided for water cooling. Insulating fiber disks $\frac{1}{16}$ inch thick are placed between layers, through which heat is conducted to the tubing. Each coil is clamped by studs between large aluminum rings $1\frac{1}{4}$ inches thick with outer diameters of 96 inches and inner diameters of $52\frac{1}{2}$ inches to fit the magnet poles, forming a compact and self-supporting unit. The two coils are mounted on the yokes with supporting studs and spaced 4 inches from the yokes to reduce the short-circuiting flux in the corners. Parallel feed of water-cooling circuits ensures adequate flow at low pressures. The coil terminals are connected in series giving an electrical circuit of approximately 2 ohms total resistance, but can be paralleled to match a lower voltage power supply if desired. The total weight of copper is about 12 tons.

The results obtained with the magnet described are portrayed by the curve of Fig. 8 between magnetic field and exciting current. Operation at 12-Mev deuterons requires 14,600 gauss and

17 kilowatts; at 15-Mev deuterons, 16,300 gauss and 28 kilowatts. The maximum field obtainable with the 90 kilowatts available extrapolates to over 19,000 gauss, allowing cyclotron operation up to 20 Mev if other limitations could be overcome. Magnet power varies with the square of the magnetic field or directly with ion energy E below 10,000 gauss and with an increasing power of E at high inductions. So there will be an economic limit on the use of high magnetic fields to obtain high energies. The use of high permeability iron is important in raising this limit.

7. RESONANT CIRCUIT AND CHAMBER

The vacuum chamber fits between the magnet poles and the dees are symmetrically located within it. Top and bottom plates or lids of iron act as extensions of the poles and reduce the magnetic gap to a minimum. The chamber must have adequate structural strength to resist distortion by the magnetic field or when under vacuum; on 42-inch diameter lids the atmospheric force is over 20,000 pounds. Side walls and all attachments must be non-magnetic to prevent distortion of the symmetrical magnetic field. For ease of assembly and accessibility for repairs the walls should be provided with several large ports including vacuum locks for removal and replacement of targets and filament. All ports and joints must have simple, positive vacuum seals. Internal surfaces should be of high electrical conductivity and gasket seals bridged by conducting material.

For large cyclotrons the method of obtaining the necessary high dee voltages has been through the use of coaxial, quarter-wave resonant dee lines. The dees are mounted on the ends of the internal conductors, which are large enough to support themselves and the dees without any insulating supports. The outer conductors serve as the vacuum wall and are joined to the chamber. Electrically, this results in physically shortened quarter-wave lines with the dees as capacity loads across the ends. To obtain the required high dee voltages at minimum power input this circuit must have a high electrical efficiency, or Q value. $Q=\omega L/R=1/\omega CR$ at resonance. The inductance and capacitance per unit length of coaxial line are fixed quantities and the

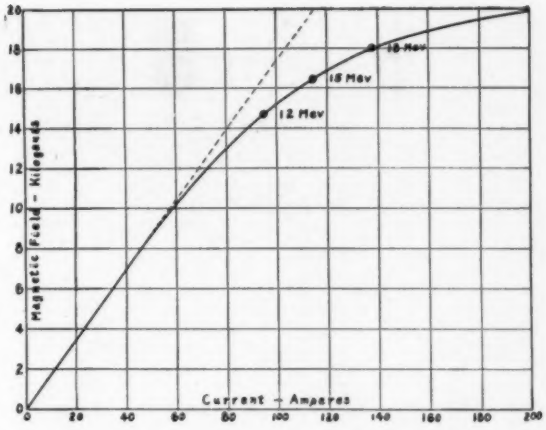


FIG. 8. Plot of magnetic field as a function of exciting current for the M.I.T. cyclotron magnet. The magnetic fields for resonance at 12-, 15-, and 18-Mev deuterium ions are indicated.

dee capacitance is essentially determined by the desire for short gap and high fields between magnet poles. High Q therefore requires low resistance and consequently large diameter coaxial lines and dee stems. The dee stem must enter the vertically narrow gap between chamber lids with adequate clearance. In earlier installations, the dee stem was reduced to a small diameter at the pole edges by a conical section. This resulted in a considerable increase in resistance at this throat constriction and consequent reduction in Q. The M.I.T. cyclotron has the round dee stems tapered down by flattening into an oval section at this point, retaining the full surface area of the large diameter stem for electrical conduction through the throat. The mechanical rigidity of this design is a valuable additional feature and the dee structure is based upon a continuation of this oval pipe. In choosing an adequate size of coaxial line, the increased resistance of this throat section and the equivalent resistance of the resonant ions must be considered. Beyond a certain point further increase in size of coaxial line is ineffective in increasing the operating value of Q. The lumped dee capacitance and the inductance of the dee and conical throat section leave only 4 to 6 feet of the cylindrical coaxial line for the operating frequency range. In the cyclotron described, the dee stems and lines were chosen to be 8 and 24 inches diameter, respectively. The calculated Q

FIG. 9. Shop assembly photograph of the M.I.T. cyclotron chamber, showing the construction of chamber and dees.

for this circuit is 10,000; in operation, the additional resistive load of resonant ions reduces the effective value to below 5000.

The dee structure is built onto the end of the 8-inch copper dee stem which is flattened into an oval section of $2\times11\frac{1}{2}$ inches. One side is cut away and extension faces are fastened on by monel screws spaced $\frac{3}{4}$ inch apart for electrical conductivity through the joint. The peripheral edge is half of a 2-inch diameter tube bent to have an outer radius of 20 inches, brazed to the oval dee stem. All surfaces are round and smooth with streamlined contours. On the dee having the internal deflector, an outer housing of 3-inch tubing encloses and shields the electrode. These and other features of the chamber and dee construction are illustrated in the photographs of Figs. 9 and 10. Water-cooling circuits consisting of $\frac{3}{8}$-inch copper tubing are silver soldered internally to the dee faces and edges, spaced an average distance of 3 inches apart; other tubes are soft-soldered to the inside surface of the 8-inch dee stem for cooling. The water tubes come out of the cyclotron at the back end of the dee lines. A 7-inch diameter steel pipe inside and supporting each 8-inch copper dee stem extends outside the dee line through a flexible vacuum seal. Large adjustment screws are used to locate dee position. The movable node which clamps between dee stem and line is made flexible to allow for this adjustment.

The chamber is designed to accommodate the oval dee stems through a back opening of full chamber width. The lines, cones, and dees are mounted as a unit on a truck on rails which can be rolled back, pulling the dees bodily out of the chamber. This greatly facilitates assembly or repairs on the dees and other internal structures. Four side plates covering large ports in the chamber wall are identical in size and interchangeable. Two side plates mount motor driven trimmer condenser plates by which dee capacitance can be varied to tune the dee circuits to frequency. The other plates are used for the probe target port and for the emergent beam port. A large front window in the diametral line of the dees sees practically the whole internal structure of the chamber and has proved invaluable for operational inspections. The ion source is mounted opposite the window on the end plate joining the two dee cones with the tubular stems fitting against top and bottom plates along the diametral line between dees, the filament stem being withdrawn through a vacuum lock.

The chamber proper consists of $1\frac{1}{4}$-inch brass top and bottom plates with large circular holes for the iron plates which form the lids. Vertical posts and side walls space the plates $4\frac{1}{2}$ inches apart internally and are silver soldered (in a furnace with silver solder foil between surfaces[16]) to form the chamber structure. The full back opening, the four 22-inch ports, and the front window leave little but a framework when opened, so essentially all repairs and adjustments can be made without removing the chamber from between the magnet poles. The Armco iron disks are 42 inches diameter and 2 inches thick, copper plated on the inside surface for electrical conductivity. They are spaced 5 inches apart, resting on $\frac{1}{4}$-inch ledges turned into the brass plates of

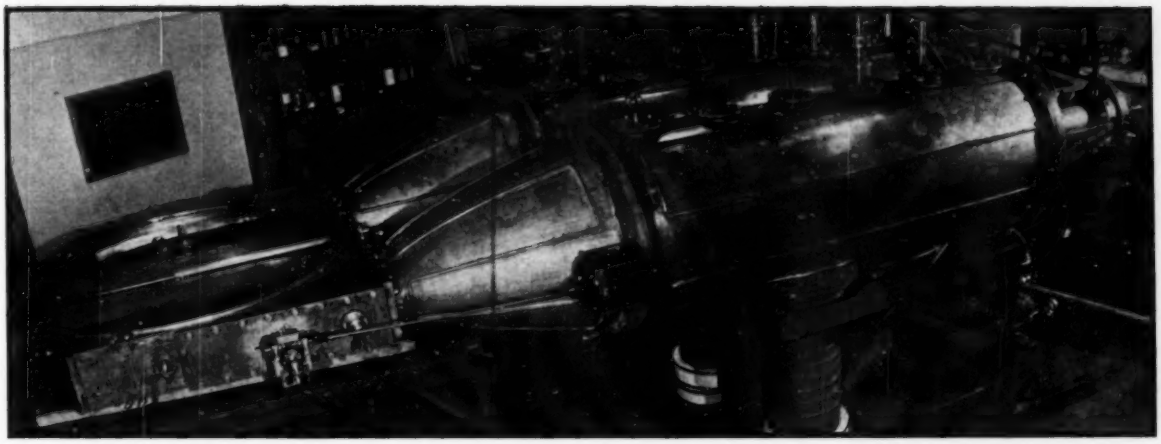


FIG. 10. Shop assembly photograph of the chamber and dee lines showing the oval throat section where the dee stems enter the chamber, the diffusion pumps, and other details.

the chamber with a packing ring around the edge for a vacuum seal (see Section 12). Shimming gaps $\frac{1}{4}$ inch wide are left between lids and pole faces. The chamber is located between poles by adjustable bolts in the lids having flat hex heads in the shimming gap.

8. ION SOURCE

The ion source generally used for the cyclotron is a development of the metallic capillary arc.[17] It is essentially a low voltage gaseous discharge within a small metallic cavity at the center of the cyclotron chamber maintained by emission from a filament. Since the magnetic field collimates the electron emission along the lines of force, the discharge is limited to a vertical column the size of the filament which crosses the cyclotron chamber. In the original design the discharge traversed a vertical copper tube (the capillary) pierced with a small hole at the central plane through which ions emerged. The hole provided flow resistance for the gas which enetered through the filament stem, thus maintaining the gas pressure of 10^{-2} to 10^{-3} mm of Hg necessary for the discharge. The arc could be struck between anode and cathode, one located above and the other below the capillary in similar housings. Or the filament housing could itself be used as the anode, the collimated discharge still travelling the length of the capillary.

With the development of high speed pumps which can handle larger gas flow, the side hole has enlarged to become a section cut out of the central portion of the tube. See Fig. 11. The tube or capillary has become a short cone with large end joining the filament housing and small end (about $\frac{1}{4}''$ inside diameter) ending just short of the central plane. The second cone is retained for symmetry of electric fields and also houses a "test anode." This test anode is a water-cooled tungsten-surfaced plate against which the vertical discharge column strikes. In the circuit a meter registers the discharge current which crosses the gap between cone tips as distinct from the total discharge current between the filament and the grounded housing, and is useful in lining up the filament with the cones.

Several modifications of this design have been used successfully in other laboratories. The

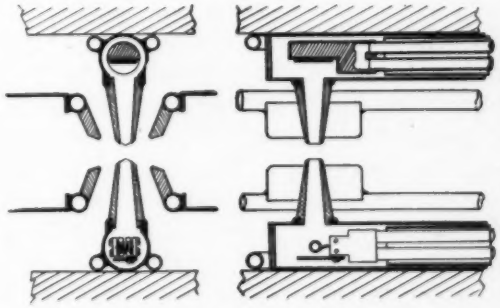


FIG. 11. Cross section diagram of the ion source showing the filament, cone shaped tips, and test anode. The filament stem is withdrawn through a sliding seal and vacuum lock for replacement.

second cone housing the test anode need not be present,[18] in which case part of the discharge column is exposed. The lack of electrical symmetry does not seem to affect operation, but the exposed half-column of discharge off the central plane tends to load the cyclotron with off-focus or semiresonant ions and more power is required to attain the necessary high dee voltages. Also the advantageous features of the test anode for adjusting filament position and measuring the effective discharge current are absent. The single-jet source has been found necessary in several cyclotrons in which provision was not made in design for a double stem mounted along the diametral line between dees. A modification of the single-stem source,[11,19] which does not have the discharge column exposed, extends the cone across the central plane as a hood with a blank end against which the discharge strikes and with an aperture in the side at the central plane facing one dee. This is feasible since the first ion paths are wide spirals which will easily clear the central hood.

The filament is a tightly wound helix of heavy tungsten wire, 50- to 80-mil diameter, heated by low frequency power (150 kilocycles) from an auxiliary 2-kilowatt vacuum-tube oscillator unit and fed by transmission line to an impedance matching unit near the filament terminals. Direct current heating usually results in magnetic distortion of the filament and shortens the useful life. Filament life is primarily a function of gas composition and is materially shortened by a trace of oxygen as from a small air or water vapor leak or contaminated deuterium gas. With good vacuum conditions 70-mil tungsten filaments have given 80 hours life. A vacuum lock is provided on the filament stem to allow changing filaments without loss of vacuum in the cyclotron chamber. The cone tips tend to erode by sputtering. Elkonite[20] (a copper-tungsten alloy) has shown the best resistance to this erosion and does not require local water cooling.

The arc source is usually operated at about 3 amperes total current, of which about 2 amperes cross the $\frac{1}{4}$-inch gap between cone tips and strike the test anode. The potential drop across the arc may vary from 80 to 150 volts depending on filament temperature and gas pressure, and is supplied by a small motor generator of 300 volts maximum output. Sloping the cone tips to favor emission toward one dee has been tried but does not seem to improve emission over a symmetrical arrangement. Extensions on the dee faces (called "feelers") increase the high frequency field in the gap between cone tips and improve ion emission. Since increase in dee voltage usually results in increased beam intensities, it is believed that the emission is space charge limited by the dense ionization in the exposed column between cone tips. Under favorable conditions resonant beam currents up to 1.0 ma have been obtained, and the ion source emission is estimated to be of the order of several milliamperes.

9. SHIMMING

The adjustment of the magnetic field for best operation by means of thin plates or shims of soft iron is known among cyclotronists as shimming. At least three purposes are served, requiring different methods, and they will be discussed separately. One purpose is to compensate for the decrease in the magnetic field near the pole edges caused by fringing and so to increase the diameter of the uniform region. This is accomplished by placing a pair of iron rings of the same external diameter as the chamber plates against these plates inside the chamber. The extension of the region of uniformity for soft iron rings having cross sections $\frac{1}{4}\times\frac{1}{2}$, $\frac{1}{4}\times\frac{3}{4}$, and $\frac{1}{4}\times 1$ inches is observable in Fig. 12 in which percent variation of the magnetic field from the value at the center (13,500 gauss in this measurement) is plotted against distance from the center.[21] The $\frac{1}{4}\times 1$ inch rings, for example, extend the region of uniform field by about 1 inch radially. This makes it possible to locate the exit slit of the deflector channel that much farther from the center and so to obtain a higher energy beam. The M.I.T. cyclotron now uses the $\frac{1}{4}\times 1$ inch rings and has the exit slit at 19-inch radius or 2 inches inside the pole edge with a 5-inch gap. The limit in the use of such ring shims is that they should not be so large as to cause a "hump" in the magnetic field (Fig. 12), which would result in a defocusing condition at that radius.

The second purpose of shimming is to provide

magnetic focusing for the ion beam while it is being accelerated. This can be accomplished by flat-pyramidal piles of thin iron disks symmetrically placed in the $\frac{1}{4}$-inch shimming gaps between magnet poles and chamber plates. The purpose is to maintain a slowly decreasing field from the center outward. This can be visualized as producing magnetic lines of force concave toward the center throughout the whole cyclotron chamber. Since the force on the ions is perpendicular to the direction of the lines of force, this will result in focusing toward the central plane. Pyramids constructed of 6 disks 0.020 inch thick of 8, 12, 16, 20, 24, and 28 inches diameter gave the result plotted in Fig. 12, an approximation of the desired continuously decreasing field.

A technique has been developed for the experimental determination of the correct size and arrangement of such shims. It uses a temporary quartz probe inserted between the dees through a sliding seal in the front window port. The quartz is a vertical strip crossing the chamber between the dee faces, so that the resonant ion beam strikes and heats the central portion to incandescence. The supporting rod can be slid in or out through the vacuum seal to place the quartz at any desired distance from the center. Starting with the quartz near the center of the chamber and withdrawing it toward the edge, the height of the heated portion and so the quality of focusing can be observed visually through the window. If the resonant ion path becomes too broad at a given radius, insertion of a pair of disks of about 1 inch smaller radius will provide the proper focusing field through this region. In this manner the minimum number of disks required for focusing can be determined. This should not increase the field in the center by more than 1 percent, in order to stay within the limits of uniformity required for resonance in the cyclotron.

A third use of shims is in correcting variations from uniform field caused by imperfections in magnetic uniformity and mechanical construction. This will be a function of the magnet involved and few general rules can be given. When the M.I.T. cyclotron was first operated the ion beam melted holes in the upper dee surfaces.

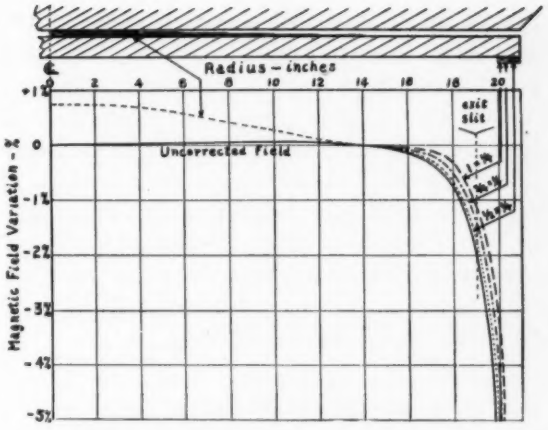


FIG. 12. Magnetic field measurements in percent of total field as a function of radial distance from the center of the chamber. The extension of the uniform region by ring shims and the radially decreasing field obtained with disk shims in the shimming gap are indicated. (Courtesy of Dr. J. R. Downing.)

This was due to a slight magnetic asymmetry and the quartz probe showed the radial location at which the beam deviated from the central plane. If the ion path rises off the central plane a disk in the upper shimming gap will produce a relative increase in field in the upper portion of the chamber and return the ion path to the central plane. Use of the quartz probe made it possible to choose asymmetrical disk shims which maintained the resonant beam close to the central plane for all radii. Other kinds of asymmetry may require spot shims in both upper and lower gaps, such as might be needed to compensate for local magnetic inhomogeneities in the chamber lids. A careful plot of magnetic field near the surfaces of top and bottom poles is useful in indicating regions of low field where such spot shims can be located. In practice a comprehensive cut-and-try procedure will usually result in a worth-while increase of beam intensity.

A common fault in an operating cyclotron is a migration of ion paths during acceleration so that the center of the last path is not at the geometrical center of the chamber. In this case, the emergent ions will not have the proper curvature for the deflector channel or the theoretical energy. This has resulted in the emergent beams having less than the theoretical energy computed from the radius of the exit slit in practically all cyclotrons. It is a natural consequence of shimming to obtain maximum beam intensity and

can do no harm except for the error in statement of ion energy. One technique which may minimize this difference is to shift the stacks of pyramid shims off center in a proper direction. Often, however, large, thin, asymmetrical shim plates in both top and bottom gaps are needed to increase the field over a large sector. It can be said in conclusion that despite widespread rumors, shimming is a relatively simple problem for the experimenter once this stage is reached, and usually has to be done only once.

10. THE DEFLECTOR

The purpose of the deflecting electrode is to pull the resonant ion beam out of its circular path and direct it against an external target. At the chosen maximum radius, R, the dee wall is cut away and a new dee wall inserted having a larger radius of curvature than the ion beam. The leading edge of this new wall is usually a thin tungsten septum, which splits the resonant ion beam so that a large fraction passes out of the dee behind the septum. A paralleling deflecting electrode with a negative d. c. potential relative to the dee wall provides an electric field to deflect the ion beam into the channel between dee and deflector. The beam diverges and straightens as the ions cross through the weakening magnetic field at the pole edges, and passes out of the chamber tangentially through a suitable port. External targets are mounted outside this port.

The d. c. potential for the deflector is provided by a high voltage transformer and kenotron supplying up to 100 kv. The voltage needed is a function of ion energy E, necessary width of channel to contain the ion beam d, and the change in radius of path r. An estimate can be obtained by combining the relations for radius of path with or without the deflecting potential P:

$$\left.\begin{aligned} Hev-\frac{Pe}{d}&=\frac{mv^2}{R+r} \\ Hev&=\frac{mv^2}{R} \end{aligned}\right\} \quad \frac{P}{d}=\frac{mv^2}{e}\left(\frac{1}{R}-\frac{1}{R+r}\right)$$

$$\text{or} \quad \frac{P}{d}=\frac{2E}{R}\left(\frac{r}{R+r}\right). \quad (10)$$

Evaluating for the 19-inch radius and 15-Mev energy of the M.I.T. cyclotron with an average deflector spacing of 0.4 inch we find:

$$P=56{,}000 \text{ volts for } r=0.1R,$$
$$P=103{,}000 \text{ volts for } r=0.2R.$$

A step-by-step calculation based upon the chosen $r=0.12R$, a variation of d along the channel from $\frac{1}{4}$ inch to $\frac{3}{4}$ inch, and measured values of magnetic field in the fringing region predicts a value of $P=57{,}000$ volts, and shows that a circular shape of the dee wall is a good approximation of the true ion beam path. The voltage required in operation to give maximum emergent beams varies from 40,000 to 60,000 depending on deflector spacing. A similar calculation extending beyond the end of the deflector is used to locate the position of the exit port in the chamber wall and the external targets. Corona discharge on the leads and voltage supply or electrical breakdown in the chamber will give a practical limit to the d. c. potential (70 kv in our case). The increase in radius of curvature r should be consistent with this limit for maximum energy ions with a suitable gap for the diverging ion beam.

In early designs the deflector was mounted through the chamber wall outside of and paralleling the dee wall, and extending from the midpoint of the dee around some 70 degrees of arc. The radiofrequency field between dee and deflector added to the d. c. field in varying degree as the ions traversed the deflector channel, depending upon the r-f phase as a function of angular position of the ions. This decreased somewhat the d.c. potentials required. Some of the problems resulting from this design were the r-f discharge initiated by d. c. sparks between dee and deflector, insulation for and by-passing of the r-f picked up by the deflector, and the variation in gap caused by mechanical vibration or thermal distortion of the dee. With the installation of large coaxial dee stems at the University of California, the method was introduced of supporting the deflector on a central pipe coaxial with the dee stem and enclosed in an outer housing attached to the dee. This results in elimination of the r-f voltage on the deflector and of some insulation and discharge difficulties

and better control of the dee-deflector spacing. The M.I.T. cyclotron has adapted this deflector design to fit the oval dee stems by preserving a 3-inch diameter channel through the throat. It is also possible to locate the septum ahead of the midpoint of the dee if desired (20 degrees in the M.I.T. design), or deflect through a larger angle than 70 degrees.

The radius R is determined by the fringing of the magnetic field and can be increased to a practical maximum by shimming as described in Section 9. The larger R can be made, the simpler is the problem of deflecting the ion beam. The tungsten septum intercepts a fraction of the resonant beam (which has a uniform radial distribution at this large radius), determined by the ratio of its thickness to the ΔR between successive ion paths. If the r-f voltage between dees is $2V$, this ΔR is given by:

$$\frac{\Delta R}{R}=\frac{1}{2}\frac{\Delta E}{E}=\frac{2V}{E} \qquad (11)$$

since $\Delta E=4V$ (two accelerations).

With $2V=140$ kv and $R=19.0$ in., $\Delta R=0.18$ in., so a 0.025-inch septum intercepts a minimum of 14 percent of the ion beam, and a larger proportion if it is not set precisely parallel to the ion path. Warping of the septum due to heat developed by this ion bombardment may cause an even larger loss. Such a tungsten septum usually gets red hot. A thin septum (0.010 inch) will warp and melt. A septum 0.025 inch thick with a slot the width of the beam cut out for several inches and with leading edge tapered and ground sharp is more successful in transmitting maximum beam. The heating and warping have been reduced in our design by silver soldering the edges of the tungsten septum to copper bars in contact with water-cooling tubes. This ΔR limitation defines a minimum dee voltage below which no emergent beam can be obtained, even though resonant ions at radius R are still available for probe targets inserted between the dees. So the fraction of the total resonant ions in the emergent beam should become greater as dee voltage is increased. It can be noted that the minimum width of the deflector channel at the entrance is determined by this ΔR and that a dee voltage which gives a larger ΔR will require a wider deflecting gap and so a higher deflecting potential.

REFERENCES

(1) E. O. Lawrence and N. E. Edlefsen, Science **72**, 376 (1930).

(2) M. S. Livingston, Ph.D. Thesis, University of California (1931).

(3) E. O. Lawrence and M. S. Livingston, Phys. Rev. **38**, 834 (1931).

(4) E. O. Lawrence and M. S. Livingston, Phys. Rev. **40**, 19 (1932).

(5) M. S. Livingston, Phys. Rev. **42**, 441 (1932).

(6) E. O. Lawrence and M. S. Livingston, Phys. Rev. **45**, 608 (1934).

(7) E. O. Lawrence and D. Cooksey, Phys. Rev. **50**, 1131 (1936).

(7a) E. O. Lawrence *et al.*, Phys. Rev. **56**, 124 (1939).

(8) M. E. Rose, Phys. Rev. **53**, 392 (1938). Equation (6) above is a composite of Rose's Eqs. (13), (15), and (11a).

(9) R. R. Wilson, Phys. Rev. **53**, 408 (1938).

(10) L. H. Thomas, Phys. Rev. **54**, 580, 588 (1938).

(11) A. Allen, M. B. Sampson, R. G. Franklin, J. Frank. Inst. **228**, 543 (1939).

(12) Columbia University Cyclotron.

(13) University of California Cyclotron. Private communications from Mr. Winfield W. Salisbury.

(14) The neutralized oscillator was designed by Mr. J. H. Cook of the M.I.T. Electrical Engineering Department as an answer to the "snap-tuning" limitation encountered in the original self-excited oscillator circuit. We are also greatly indebted to Mr. Winfield W. Salisbury for many helpful discussions and suggestions during the course of this development.

(15) Professor H. B. Dwight of the M.I.T. Electrical Engineering Department confirmed the Armco design by an independent calculation of the magnetic circuit and the tapering of the poles.

(16) An error in furnace technique caused the brass plates to sag. On being straightened by cold working some of the silver solder joints cracked open and were inadequately reworked. The result has been some trouble with vacuum leaks at the repaired joints, but the undisturbed joints have been sound and the error could be avoided. Several satisfactory alternative designs for the chamber are discussed in Section 12.

(17) M. S. Livingston, M. G. Holloway, and C. P. Baker, Rev. Sci. Inst. **10**, 63 (1939).

(18) E. McMillan and W. W. Salisbury, Phys. Rev. **56**, 836 (1939).

(19) R. L. Thornton, Washington University, St. Louis, Missouri, private communication.

(20) B. R. Curtis and R. S. Bender, Rev. Sci. Inst. **13**, 266 (1942).

(21) Observations by Dr. J. R. Downing.

Here and There

Staff Changes

Hubert M. Turner of New Haven has been elected President of the Institute of Radio Engineers for the coming year, it has been announced by the Board of Directors. He is Associate Professor of Electrical Engineering at Yale University and succeeds Dr. Lynde P. Wheeler of the Federal Communications Commission, Washington.

A 38-year-old engineer who developed some of the newest secret equipment used in military radio communications has been appointed Associate Director of the Westinghouse Research Laboratories, it was announced recently by Dr. L. W. Chubb, Director of the Laboratories. He is John A. Hutcheson, for the last three years Manager of Engineering at the Baltimore Radio Division of the Westinghouse Electric and Manufacturing Company. In that post he directed a large staff of engineers developing new types of radio equipment for the armed forces and designing secret electronic devices for military use.

R. H. Siemens has been appointed Chief Engineer of RCA Victor Argentina, wholly-owned RCA subsidiary company in Buenos Aires, it has been announced by J. D. Cook, Managing Director of RCA Victor's International Division.

Mathematical Tables Project

Four tables, the work of the Mathematical Tables Project, conducted under the sponsorship of the National Bureau of Standards, are being published by the Columbia University Press. The first three tables were to be ready for publication about the middle of November, and the fourth is expected to be published early in 1944. Orders for the tables may be placed with the Columbia University Press, Box E819, 2960 Broadway, New York 27, New York.

New Booklets

A comprehensive practical manual on industrial salvage has just been published by the Technical Service Section, Industrial Salvage Branch, Salvage Division, War Production Board, and is now being distributed to industry. The price is $.50 per copy and copies are procurable through the Superintendent of Documents, Government Printing Office, Washington, D. C. Excerpts are permitted provided proper credit is given.

The Spectrographer's News Letter, No. 2, is a bi-monthly paper published by Harry W. Dietert Company, 9330 Roselawn Avenue, Detroit, Michigan, and can be obtained by writing to this address with a request to be put on the company's mailing list.

Slidefilms and Motion Pictures to Help Instructors, an 84-page catalog published by the Jam Handy Organization, was prepared to help the teacher select discussional slidefilms and motion pictures for use in teaching. Copies may be obtained from the Jam Handy Organization, 2900 East Grand Boulevard, Detroit 11, Michigan.

Bakelite Review, October, 1943, is a quarterly publication of Bakelite Corporation and the Plastics Division of Carbide and Carbon Chemicals Corporation, 30 East 42 Street, New York 17, New York. Its purpose is to inform industry of the latest developments in plastics and their application.

Reprints from the *Journal of Mathematics and Physics*, May and June, 1943, on New Mathematical Tables, can be obtained from the National Bureau of Standards, Information Section, Washington, D. C. The tables referred to are MT20, MT21, and MT22.

Recent Applications of Physics

Possible Applications of Ultra-Short Radio Waves Ultra-short radio waves which today are beamed like light to locate distant enemy aircraft and perform other miracles of war, will be used after the war to perform equally striking feats in the home, on the highways and high seas, and in the factory, according to I. E. Mouromtseff, electronics engineer and pioneer experimenter with short wave, at the Westinghouse Electric and Manufacturing Company.

Small ultra-short wave radio sets, similar to military equipment, might be installed on automobiles. When these radio beams strike rows of overhead or roadside reflectors, the driver would be warned of danger spots or could be guided along the road despite the densest fog. No longer will one train crash into the back of another on the same track nor will ships have difficulty finding the way into and out of harbors in fog and rain. Long wave radio gave us sound at a distance; the very short waves now give us vision at a distance. Long waves, like those used in standard broadcasting, radiate from a sending antenna; they travel in all directions and use the earth's surface as a guide, thus following a slightly curved path. Ultra-short waves, on the other hand, are "channelized" into a narrow path and travel through the air in a straight line. Like light, these short radio waves can be projected in a beam and reflected.

Among the "feasible applications" in industry for this form of radio are such jobs as helping to convert chemicals into cloth, aiding in the manufacture of safety glass, and treating preserved foods after they already are in the jars or packages. One of the first industrial experiments with ultra-high frequency radio was for the de-infestation of wheat in grain elevators contaminated with rice weevil. Also vacuum packers have become interested in the possibility of precooking hams and similar products by high frequency radio. Restaurants have investigated the future of large "radio cookers" which might be installed in dining rooms for grilling steaks, or toasting bread in front of the customer. Heat would be generated only in the food itself and the food would be cooked from the inside out giving it a different flavor. In the field of medicine it has been found that various sicknesses can be treated by irradiating the body with radio waves. Very short waves can be focused into narrow beams—narrowcast instead of broadcast—and applied to the body for therapeutic treatment at sharply defined places. Radio, for example, can be used for irradiation of kidneys, lungs, stomach, and other areas. It is also the most convenient method of heating the entire body.

Stronger Tin Alloys A further contribution to the development of stronger tin base alloys suitable for use as bearing metals is reported in a paper by D. W. T. Pell-Walpole, in the *Journal of the Institute of Metals* for June, 1943.

The tensile properties of a series of tin-antimony-cadmium alloys were examined after quenching and prolonged tempering treatments. It has already been shown (*Journal of the Institute of Metals*, October, 1943) that the hardness of tin-base alloys in the range antimony 9 to 10 percent, cadmium 1 to 1½ percent, balance tin, can be appreciably improved by heat treatment. It is now shown that the same range of alloys has improved tensile strength after quenching from the highest practicable annealing temperature followed by prolonged tempering at 140°C. The best alloy from this point of view contains 9 percent antimony, 1½ percent cadmium, balance tin. The tensile strength of the alloy reaches 6 tons per square inch as quenched and tempered, and this strength is retained at 140°C. This temperature was chosen as being near the maximum at which tin-base bearings may have to operate continuously.

Reprints of Dr. Pell-Walpole's paper (Publication 114.A) may be obtained on request to the Tin Research Institute, Fraser Road, Greenford, Middlesex.

G-E Vibration-Velocity Meter A General Electric electronic vibration-velocity meter was used recently to detect the cause of vibration in grinding machines at the plant of John Bath & Company, Worcester, Massachusetts, manufacturers of high precision taps. The vibration in one machine in particular was causing a large number of rejects. An investigation made with the aid of the instrument revealed that a set of gears in this machine, apparently in good condition, caused the vibration. Replacing the gears eliminated the difficulty.

The Bath concern also found the meter valuable for detecting low vibration areas in its plant, thus facilitating the placing of new equipment in areas where maximum operating efficiency was assured. Consisting of a vibration pick-up unit and electronic amplifier unit, the meter measures vibration velocity and, together with an integrating unit, vibration displacement. The amount of vibration can be analyzed graphically by the use of an oscilloscope fed by the amplifier.

Dark Green Paint Camouflage A dark green paint is now being made by the Du Pont Company. This new camouflage finish reflects the infra-red rays that, along with ultraviolet, represent the unseen three-fifths of "sun's spectrum." It shows up light instead of dark in aerial infra-red photographs. In contrast, ordinary green paints appear dark in aerial photographs, revealing details to the enemy.

Detection of camouflage depends chiefly on visual observation and lens work aloft. The military value of infra-red photography is based on the fact that common objects reflect visible light and infra-red rays in a quite different manner. Most green vegetation, for example, reflects infra-red light. A military target blended into the landscape by the artful use of orthodox green paint would not escape detection by the infra-red aerial camera, because the painted surfaces would stand out in marked contrast to the foliage, black against white in the photograph.

Now, however, the camoufleur is one jump ahead of the recently developed infra-red technique. He may use green paints that not only visibly match surrounding natural greens, but also equal their infra-red reflectance powers and accordingly match in the infra-red photograph as well. Other colors—olive and field drab, loam black, earth red, brown, and yellow—are also possible in the range of paints that are more than they appear to be to the human eye.

Innovations in Instruments

Two Table-Type Continuous Printers

Two table-type continuous printers have been developed by Peck and Harvey, Chicago. These B-1 and B-2 models produce clear, sharp blueprints or direct process black and white prints up to 44" wide in any lengths at a speed up to 42" per minute. The printers are compact, easily portable, and require no special wiring. For complete details, B-1 and B-2 bulletin can be obtained from Peck and Harvey, Manufacturers, 4327 Addison Street, Chicago, Illinois.

Contributed Original Research

Effect of Strain Rate Upon Plastic Flow of Steel*

C. Zener and J. H. Hollomon
Research Division, Watertown Arsenal Laboratory, Watertown, Massachusetts
(Received October 7, 1943)

An experiment has been designed to check a previously proposed equivalence of the effects of changes in strain rate and in temperature upon the stress-strain relation in metals. It is found that this equivalence is valid for the typical steels investigated. The behavior of these steels at very high rates of deformation may, therefore, be obtained by tests at moderate rates of deformation performed at low temperatures. The results of such tests are described. Aside from changing the isothermal stress-strain relation, an increase of strain rate tends to change the conditions from isothermal to adiabatic. It is found that at low temperatures, the adiabatic stress-strain relation in the plastic range is radically different from the isothermal, having an initial negative rather than a positive slope. This initial negative slope renders unstable homogeneous plastic deformation.

IN most engineering structures, steels are stressed only within their elastic range. In such applications, the relation between stress and strain is independent of the rate of strain, aside from the slight difference between the isothermal and adiabatic elastic moduli. In other engineering structures the steel undergoes plastic deformation in service, and frequently this deformation occurs at a high rate of strain. In order to be able to predetermine the behavior of such structures,it is, therefore, necessary to know the relation between stress and strain in the plastic region, and the dependence of this relation upon rate of strain.

Any attempt to measure directly the stress in a specimen subject to a high strain rate must necessarily encounter uncertainties associated with the inertia of the test specimen itself. The authors[1] have recently pointed out that these uncertainties are of such a magnitude as to prevent the direct determination of the stress-strain relation at very high strain rates. They[2] later suggested an indirect method for determining this relation, a method based upon a supposed quantitative equivalence in the effects of changes of strain rate to changes in temperature. In the present paper, an experimental vindication of this indirect method is presented, examples are given of its use, and finally certain conclusions are drawn as to the effect of strain rate upon the stress-strain relation in steels at high rates of deformation. The implication of these conclusions with respect to the conditions and types of fracture is discussed in another paper.[3]

1. THEORY OF METHOD

Strain rate affects the stress-strain relation in two distinct manners. On the one hand, the speed at which the deformation occurs, together with the geometry of the specimen, determines whether the deformation is isothermal or adiabatic. On the other hand, strain rate has an effect *per se*, which is manifested by changes in the isothermal stress-strain relation.

The method used in this paper to determine

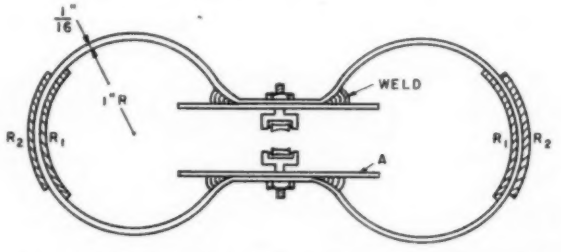


FIG. 1. Diameter gauge. R_1 and R_2—strain gauges (see Fig. 4). A—provides a place to grip the gauge when placing it on specimen.

* The statements or opinions in this article are those of the authors and do not necessarily express the views of the Ordnance Department.

[1] J. H. Hollomon and C. Zener, "High speed testing of mild steel," Am. Soc. Metals, in print.

[2] C. Zener and J. H. Hollomon, "Plastic flow and rupture in metals," Am. Soc. Metals, in print.

[3] J. H. Hollomon and C. Zener, "Conditions for fracture of steel," Am. Inst. Min. Met. Eng., in print.

the isothermal stress-strain relation is based upon a suggestion made in a previous publication that at room temperature and below, the isothermal stress-strain relation $S(\epsilon)$ in steels depends upon strain rate $\dot{\epsilon}$ and upon temperature T, only through a single parameter P:

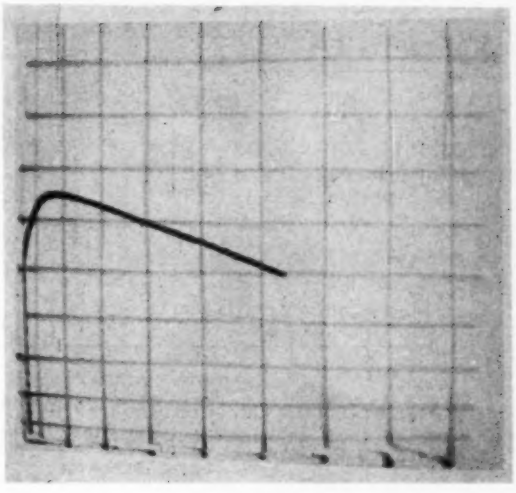
$$S=S(P, \epsilon) \tag{1}$$

and that this parameter has the form:

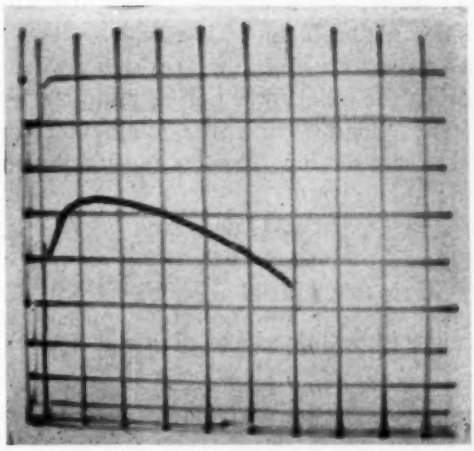
$$P=\dot{\epsilon}e^{Q/RT}. \tag{2}$$

This relation was based upon the consideration that the mechanical properties must depend only upon a dimensionless quantity, and, therefore, upon the ratio of the strain rate and some rate characteristic of the material itself. Most rates characteristic of metals, such as diffusion rates, relaxation rates, etc., are associated with heats of activation. In the writing of Eqs. (1) and (2), the assumption has therefore been made that at room temperature and below only one type of rate, such as a rate of relaxation, affects the isothermal stress-strain relation, and that this rate has a heat of activation.

The usefulness of Eqs. (1) and (2), if true, lies in the fact that the effect of a large increase in strain rate upon $S(\epsilon)$ may be obtained experimentally merely by lowering the temperature. The adiabatic stress-strain relations may then be obtained by computation. The validity of Eqs. (1) and (2) must be established by making a determination of the stress-strain relation over a wide range of strain rates and of temperatures, thereby establishing whether or not the stress-strain relation is a function of $\dot{\epsilon}$ and of T only through the single parameter P.



(a)



(b)

FIG. 2. Effect of lowering temperature on load diameter curve. (a) Strain rate $=18\times10^{-4}$ sec.$^{-1}$, temperature 20°C. (b) Strain rate $=84\times10^{-4}$ sec.$^{-1}$, temperature -130°C.

2. METHOD OF MEASUREMENT

It is necessary during the tensile tests at the several temperatures and strain rates to measure simultaneously the quantities necessary for the determination of the stress and of the strain.

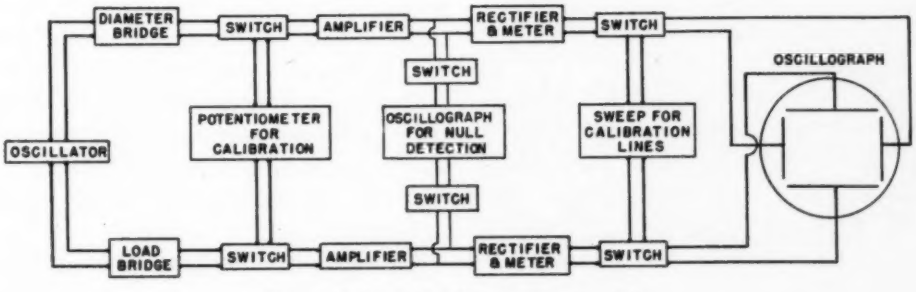


FIG. 3. Schematic diagram of assembly.

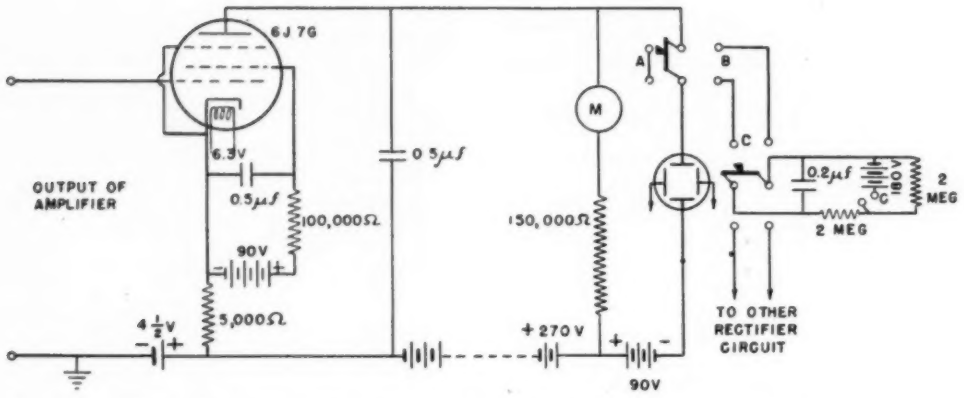


FIG. 4. Rectifier and sweep circuits. *A*—switch position during test; *B*—switch position during calibration; *C*—switch for calibration lines.

Further, these measurements must be made automatically, since at the higher speeds the test is completed in several seconds.

The stress and strain are determined by measuring the axial force acting upon the specimen (the "load") and the diameter at the smallest cross section. Then at the smallest cross section:

$$\text{Stress} = \text{Load}/\text{area}$$

and[2]

$$\text{Strain} = \ln(\text{initial area}/\text{final area}).$$

Changes in both load and diameter are related to changes in resistance of electric strain gauges attached to certain elastic mechanical elements. The elastic element for measurement of load is simply a cantilever beam deflected by the weighing arm of the tensile test machine (lever-type Riehle machine). The elastic element for measurement of changes in diameter is illustrated in Fig. 1. A tensile specimen, previously reduced slightly at the center of the gauge by machining, is placed between the two rollers. The rollers insure that the element remains at the position of minimum cross section even after the specimen begins to neck.

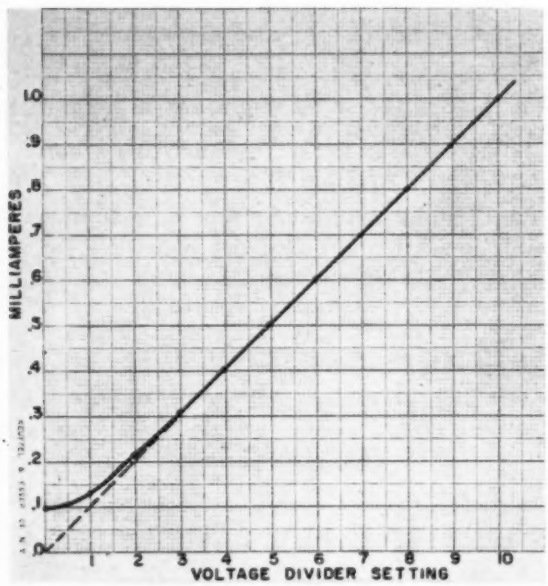


FIG. 5. Diameter gauge; calibration curve.

The change in resistance of each electric strain gauge gives rise to an unbalance in a Wheatstone bridge. A 1000-cycle current is used to excite the bridge. The alternating unbalance e.m.f. is amplified and then rectified. The rectified current is made to pass through both a milliammeter and a resistance. The milliammeter is used for calibration purposes, and for very slow tests. A pair of plates of an oscilloscope is connected directly across the resistance, so that changes in resistance reflect themselves in deflection of the oscilloscope beam. The vertical plates are connected across the resistance associated with the load bridge, the horizontal plates are connected across the resistance associated with the diameter bridge. A time-exposure photograph is taken of the oscilloscope screen, thereby giving a load-diameter curve. Typical examples are reproduced as Fig. 2.

Calibration coordinates are placed directly upon each film by means of a sweep circuit. The

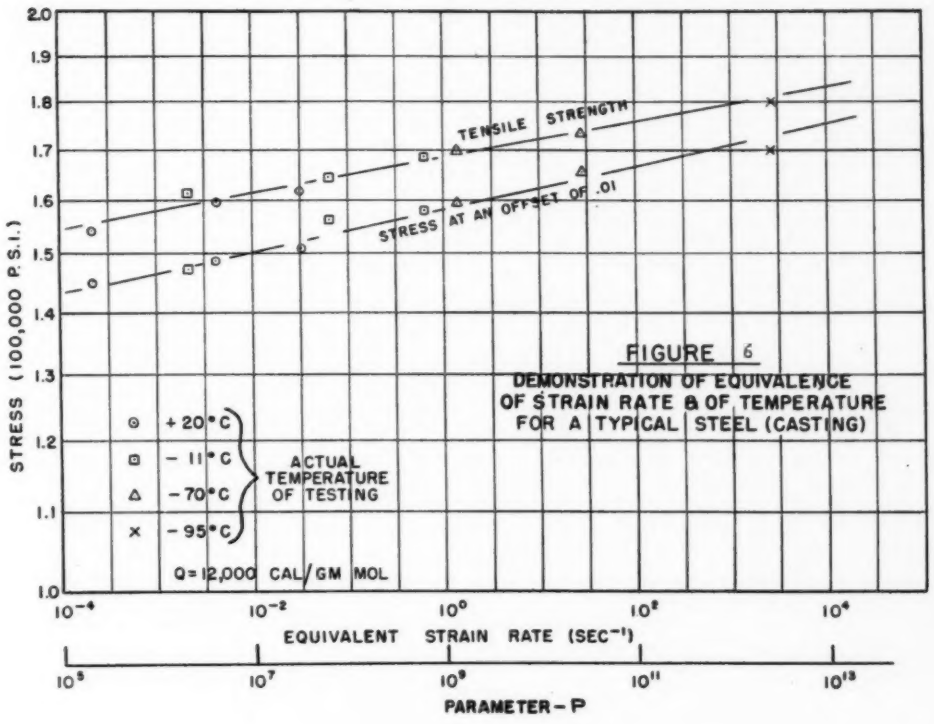


FIGURE 6
DEMONSTRATION OF EQUIVALENCE OF STRAIN RATE & OF TEMPERATURE FOR A TYPICAL STEEL (CASTING)

FIG. 6.

load bridge is balanced at zero load. A load somewhat greater than the maximum load expected is then placed on the machine, and the resulting unbalance in the corresponding bridge is divided into ten equal parts by a potentiometer. At each setting of the potentiometer the sweep circuit traces a horizontal line. Similarly, each horizontal spacing is made to correspond to a given change in diameter of the specimen.

A schematic diagram of the complete assembly is presented as Fig. 3. The only element of the electrical assembly which is not straightforward is the rectifier unit. Convenience of operation renders it necessary that the rectified impulse be linear over as wide a range as possible. Reference to electronic texts gave no indication of a philosophy of the linearity of rectifier units. The circuit shown in Fig. 4, with the indicated constants found by trial and error, was found to be fairly satisfactory, giving the calibration curve shown in Fig. 5.

The rate of strain was taken, somewhat arbitrarily, to be equal to the original gauge length divided by the speed of the machine head. Actually the strain rate increases during the test once necking begins, i.e., once the load begins to fall. However, for the present experiments, where only accurate comparative strain rates are necessary, this method is regarded as satisfactory.

For temperatures down to that of dry ice, the specimen and diameter measuring element were immersed in a constant temperature bath. For temperatures below this, the electrical insulation on the gauges became too stiff to allow sufficient flexibility. For lower temperatures (below −70°C), the two end grips were immersed in liquid nitrogen, and the temperature of the test specimen at the time of the test was measured by a thermocouple spot-welded onto the specimen near the reduced region.

The specimens were machined in the form of standard[4] 0.357″ diameter (0.1 square inch cross section) tensile bars. They were then reduced at center to 0.325″ diameter in order to insure that the necking would occur near the center. The radius of curvature of this final reduction was about 1″.

[4] *Metals Handbook* (Cleveland, 1939), p. 132.

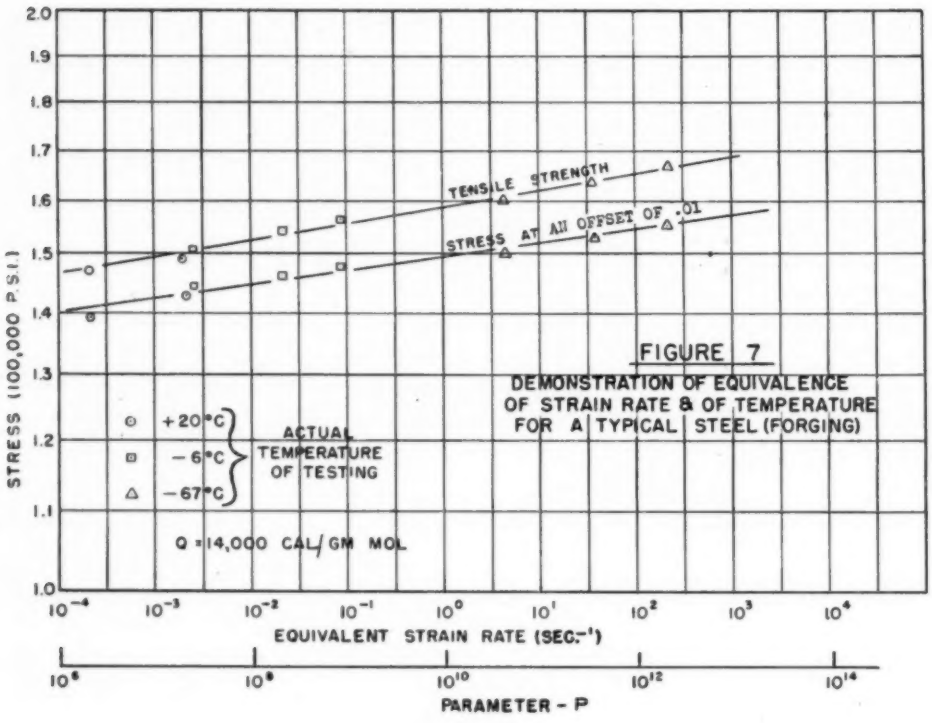


FIGURE 7

DEMONSTRATION OF EQUIVALENCE OF STRAIN RATE & OF TEMPERATURE FOR A TYPICAL STEEL (FORGING)

FIG. 7.

3. DEMONSTRATION OF EQUIVALENCE OF THE EFFECTS OF STRAIN RATE AND TEMPERATURE

The approximate composition of the steel used to demonstrate the equivalence is given in Table I. A set of tensile specimens was machined from both a large casting and a large forging. The casting and forging were heated in large sections at 900°C for six hours. This treatment transformed the ferrite into austenite and also forced all of the carbon into solid solution, thereby removing most of the effects of the past history of the steel. The large sections were water quenched, the centers of the sections thereby transforming to bainite, which consists of spheroids embedded in an acicular matrix of ferrite. Following this quenching, the steels were tempered to produce about 100,000 p.s.i. yield strength. Tensile specimens were machined from near the centers of the sections and tested at several rates and temperatures.

TABLE I. Composition (percent).

C	S	P	Si	Cr	Mo	V
0.25	0.02	0.01	0.4	0.5	0.5	0.1

The comparison of the effects of changes in strain rate and changes in temperature were made by comparing the tensile strengths (maximum load divided by the original area) and the stress at strain of 0.01. If strain rate and temperature affect the stress-strain relation only through the parameter P, then when all the stresses at a

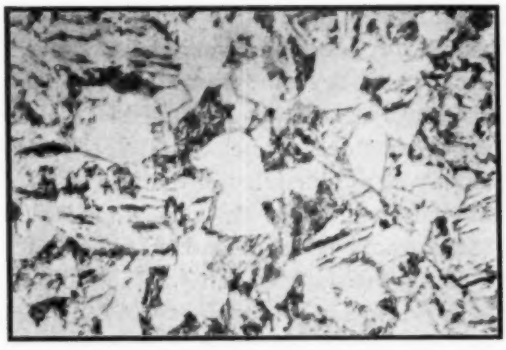


FIG. 8. Tempered pearlite used in obtaining data of Fig. 9. ×500.

given strain are plotted against P, they should lie upon a common curve. Such a plot is impossible before the constant Q is found. This constant is first fixed by finding the value which makes the observations at 20°C lie upon the same curve as the observations at the next lowest temperature. The observations made at the various temperatures and strain rates are presented in Figs. 6 and 7. The data for the tensile strength and the stress at a strain of 0.01 for both the casting and the forging, illustrate the equivalence between the effects of strain rate and temperature, for in each case the data are found to lie upon a single curve. Within the experimental error, the value of Q is independent of the strain at which the stress is measured, but it is very likely to be characteristic of the material.

4. ISOTHERMAL STRESS-STRAIN RELATION

Since most of the previous investigators of the effect of low temperature on the mechanical properties of steels used material so heat treated as to produce a pearlitic structure, specimens were machined from a casting of similar composition to that given in Table I. These specimens were heated to 900°C for one hour to form homogeneous austenite and then air cooled. This treatment results in a structure consisting of a mixture of ferrite and pearlite, the latter being an eutectoid mixture of lamellae of carbide and ferrite. Following the air cooling, the specimens were tempered at 690°C for three hours in order to produce the desired strength and ductility. This treatment partially spheroidizes the lamellae of carbide. A photomicrograph of the final material is presented as Fig. 8. The white regions represent the ferrite which precipitated out of the austenite before the latter reached the eutectoid composition (0.8 percent C).

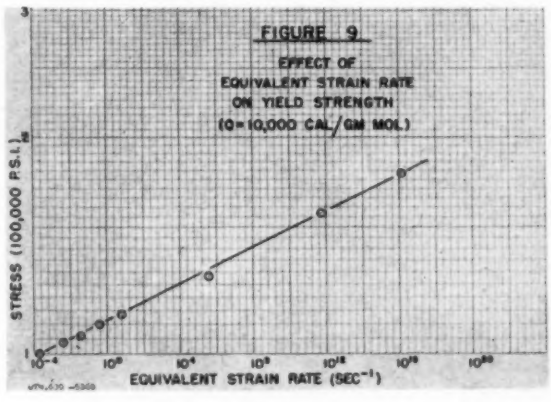


FIGURE 9
EFFECT OF EQUIVALENT STRAIN RATE ON YIELD STRENGTH (Q=10,000 CAL/GM MOL)
STRESS (100,000 P.S.I.)
EQUIVALENT STRAIN RATE (SEC⁻¹)

FIG. 9.

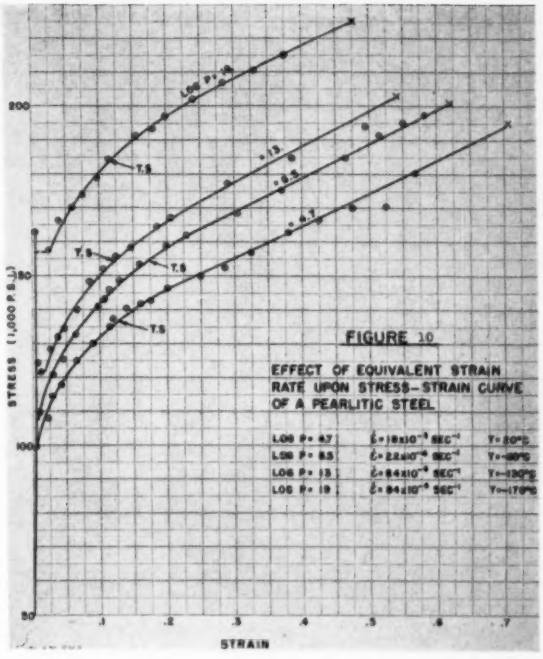


FIG. 10.

The results of tests on these specimens are presented in Fig. 9. For this material, the temperatures used varied from +20°C to −190°C. In this plot the abscissa, called the "equivalent strain rate," is the strain rate at which the material would have to be deformed at room temperature in order that the P have the same value as in the actual experiments. Typical stress-strain curves for the pearlitic steel are given as Fig. 10. These were calculated from photographic records similar to those shown in Fig. 2. The curves become straight lines beyond a strain of about 0.2, in agreement with past observations.[5–7] As may be inferred, the steel

[5] E. W. MacGregor, Trans. A.I.M.E. **124**, 208 (1937).
[6] M. Gensamer, E. B. Pearsall, and G. V. Smith, Trans. A.S.M. **28**, 380 (1940).
[7] J. H. Hollomon, "Effect of heat treatment and carbon content on the work hardening characteristics of several steels," Am. Soc. Metals, in print.

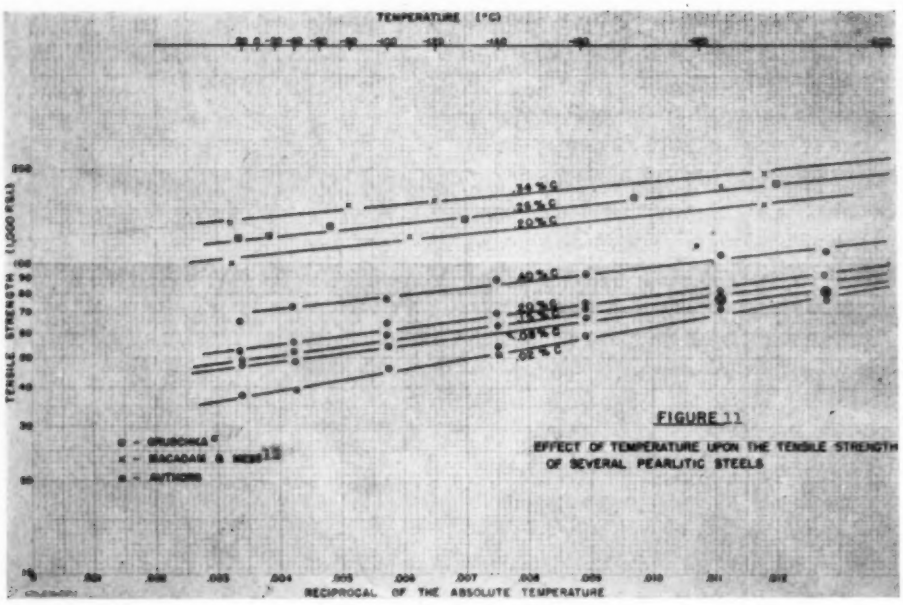

TEMPERATURE (°C)
TENSILE STRENGTH (1,000 PSI)
RECIPROCAL OF THE ABSOLUTE TEMPERATURE
FIGURE 11
EFFECT OF TEMPERATURE UPON THE TENSILE STRENGTH OF SEVERAL PEARLITIC STEELS

FIG. 11.

showed a drop in stress ("drop in load") at the inception of plastic yielding at the two lowest test temperatures. The crosses indicate the point of fracture. As shown in Figs. 6, 7, and 9 for all of the materials tested, the logarithm of the stress is a linear function of the logarithm of P:

$$S \sim P^r. \tag{3}$$

From the slope of the curve for the tempered pearlitic steel, the value of r is determined and

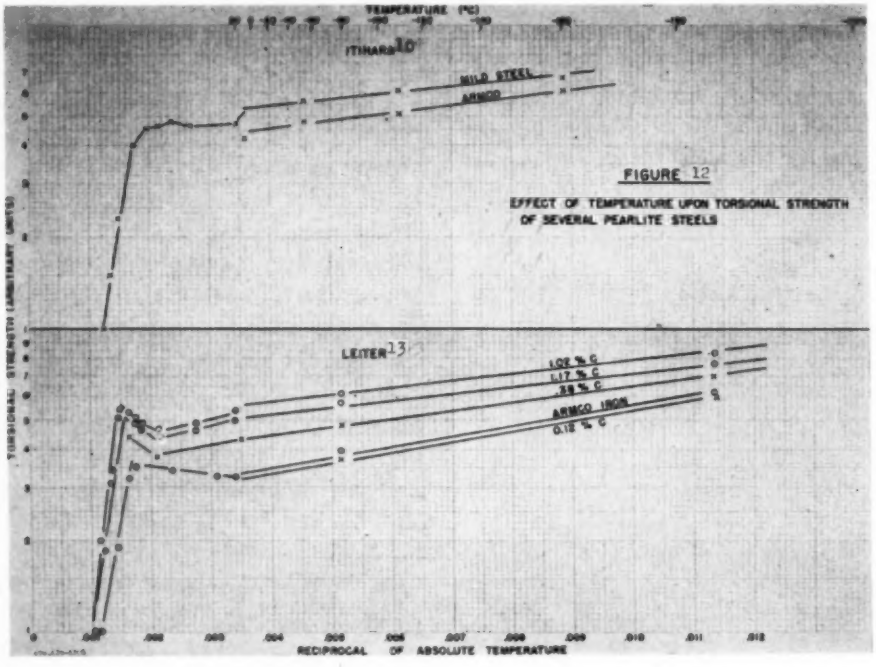


FIG. 12.

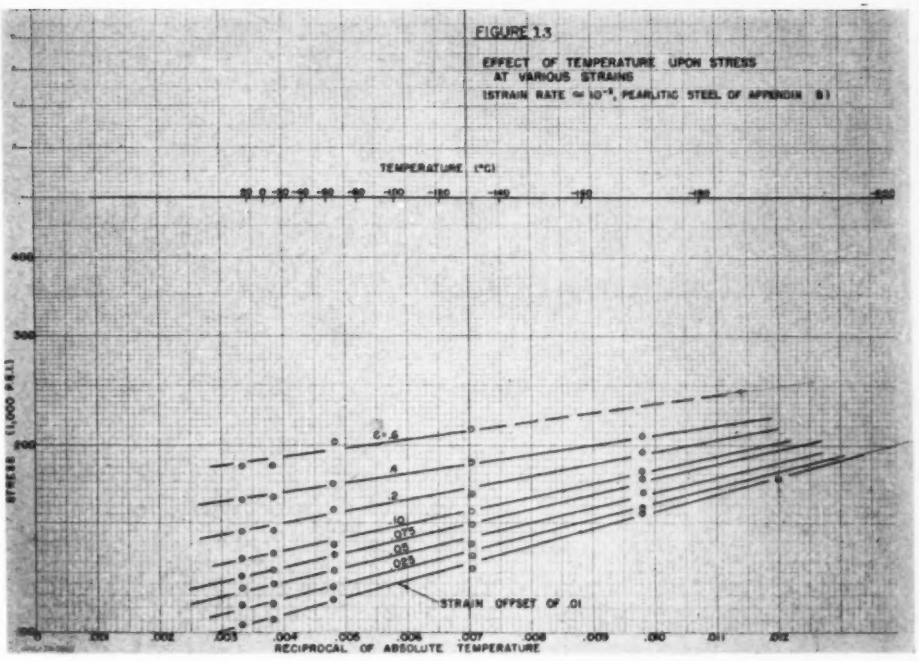


FIG. 13.

is equal to 0.0125. The power relation of Eq. (3) is in agreement with experiments of Deutler[8] upon mild steel at very low strain rates. The characteristic exponent for this material was 0.020.

The variation of stress with strain rate as given by Eq. (3) with $r=0.0125$ is extremely

[8] H. Deutler, Physik. Zeits. **33**, 247 (1932).

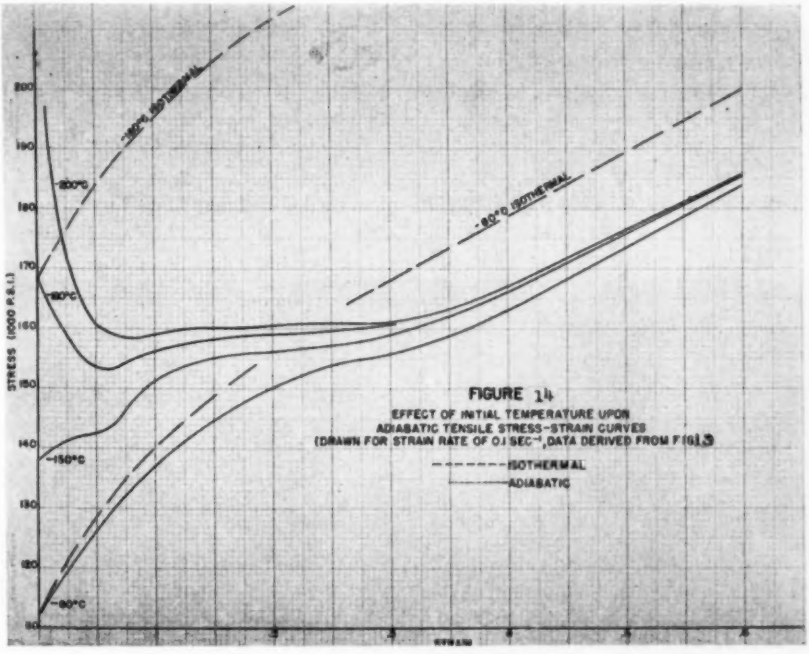


FIG. 14.

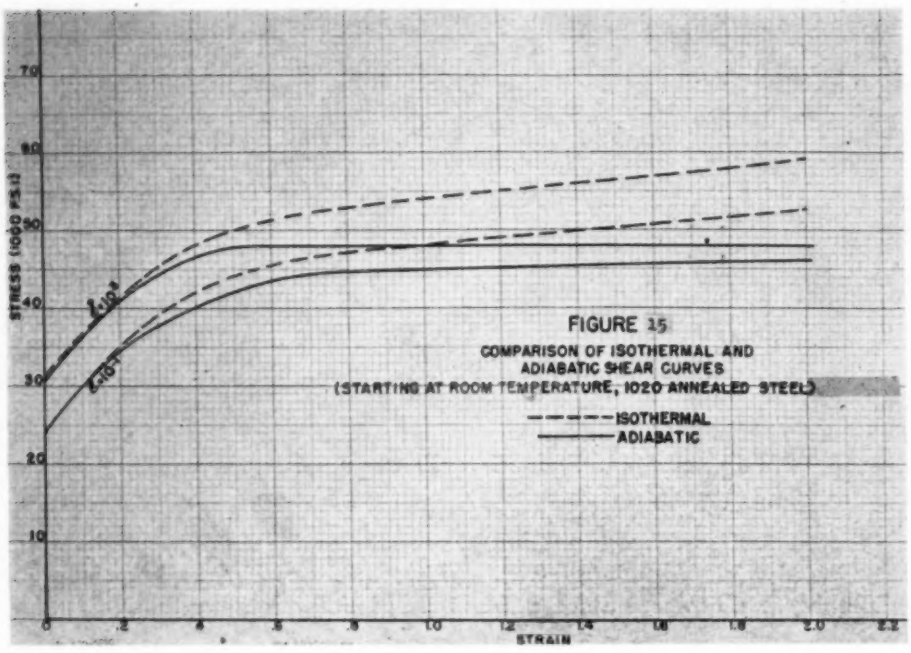


FIG. 15.

slow. Thus, to reduce the stress by 25 percent from its value when the head of the tensile machine moves at the rate of 1″ per minute, the head would have to move only 1″ in about 100,000 years. Correspondingly, if the stress were to be increased by 25 percent, the speed of the head would have to be nearly that of light.

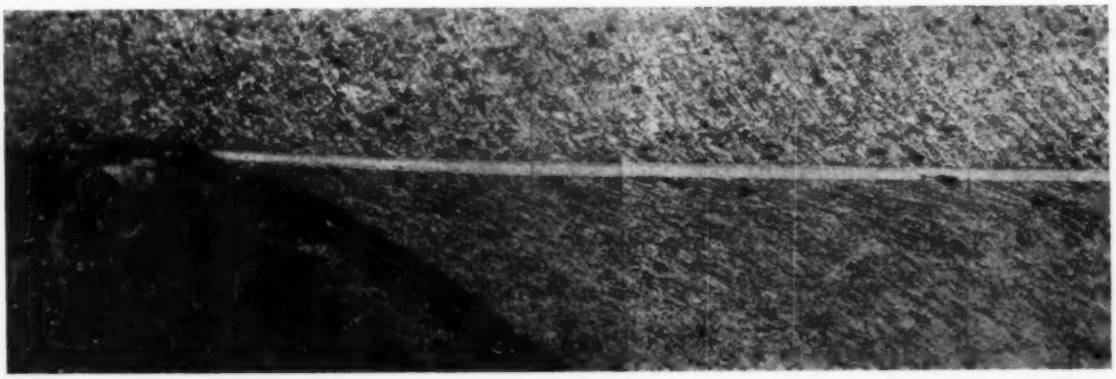
Equation (3), combined with Eq. (2), implies that:

$$S \sim e^{q/RT}, \tag{4}$$

where $q=rQ$. In order to test this relation, a plot was made of the logarithm of tensile strength (maximum load/original cross section) *vs.* the reciprocal of the absolute temperature for all the data found in the literature.[9-12] These plots, pre-

[9] M. Itahara, Tech. Rep. Tohoku Imp. Univ. **11**, 16, 512 (1933); **12**, 63 (1936).
[10] G. Gruschka, Forshung. **364**, suppl. to Forsch. a.d. Gebiete d. Ingenieur wesens, Series B, No. 5. V.D.I. (Verlag, Berlin, 1934).
[11] D. J. McAdam and R. W. Mebs, A.S.T.M. (1943) Preprint No. 40.
[12] R. W. Leiter, Thesis, Graduate School of Engineering, Harvard University (1934).

Plate



Punch

FIG. 16. Martensitic layer produced by blunt edged punch

sented as Fig. 11, confirm the relation of Eq. (4). The "heat of activation" q is nearly the same for all steels with carbon content greater than 0.08 percent, ranging from 90 to 120 calories/gram mole.

It is to be expected that the shearing stress, as measured in torsion experiments, will show the same temperature dependence as the tensile stress. All the data found in the literature are plotted in Fig. 12, and are seen to lie upon straight lines below room temperature. The slopes of these lines are very nearly equal, and lie within the range of slopes of the corresponding tensile curves in Fig. 11.

In order to show the detailed dependence of temperature, and hence of strain rate, upon the whole of the stress-strain curve, the stress at various strains *vs.* temperature is plotted in Fig. 13. The "heat of activation" q is seen to vary from 70 to 130 calories/gram mole as the plastic strain changes from 0.01 to 0.6.

5. ADIABATIC STRESS-STRAIN RELATION

Strain rate has been seen to have only a very slight effect upon the isothermal stress-strain relation. However, a change from isothermal to adiabatic deformation may be caused by an increase in strain rate. In order completely to describe the effect of strain rate, it is, therefore, necessary to investigate the case where the conditions are adiabatic.

When the material undergoes an element of strain $d\epsilon$ adiabatically, the stress is raised by the strain hardening, and lowered by the associated rise in temperature. The rise due to strain hardening may be obtained from the isothermal stress-strain curve at the pertinent temperature. The drop may be obtained from the known temperature dependence, and from the associated rise in temperature. An approximate estimate of the temperature rise may be obtained by assuming that all the work of plastic deformation is converted into heat. Actually about 15 percent is retained as latent energy.[13] By this method, adiabatic curves have been constructed for the steel represented in Fig. 13. The tensile curves starting at various temperatures are reproduced as Fig. 14.

At low temperatures, the adiabatic curves are seen to be radically different from the isothermal curves. They show an initial drop instead of a continuous rise. The drop, in stress due to the rise in temperature more than compensates for the rise in stress due to strain hardening. Further, while the separation of the isothermal curves is nearly independent of strain, the adiabatic curves nearly converge beyond a strain of 0.3.

The effect of passing from a slow isothermal to a rapid adiabatic stress-strain curve is illustrated in Fig. 15 for the case of shear. The data for the slow isothermal stress-strain curves have been calculated. It is to be noted that the high speed adiabatic curve is flat. This is in agreement with impact torsion experiments.[9]

[13] G. I. Taylor and H. Quinney, Proc. Roy. Soc. **143** 307 (1934).

Plug
on a ¼" plate of 363 Brinell hardness (nital etch, ×50).

6. DYNAMICAL CONSEQUENCES OF ADIABATIC RELATIONS

A negative stress-strain curve implies an intrinsic instability of the material where deformation cannot be homogeneous; for a region which, by chance, suffers more deformation than the surrounding region and becomes weakened thereby, will continue to deform while the surrounding region undergoes no further strain. Examples of instability are well known. In mild steel which has a "drop in load" at the initiation of yielding, plastic deformation is at first confined to localized regions called Lüders' bands. After the maximum load is reached in a tensile test, deformation is confined to a localized region, the "neck."

Reference to Fig. 14 shows that in steels at low temperatures, initial deformation must be heterogeneous at rates of strain sufficiently high to insure adiabaticity. Under these conditions, a specimen in tension will initially contract locally before the rest of the gauge length suffers any deformation; a specimen in compression will bulge locally before the remaining portion suffers any plastic compression.

The authors have obtained such an instability by using a standard type die to form a punching in a plate. A very high strain rate, about 2000 sec.$^{-1}$, was obtained by dropping a weight upon the punch, thereby attaining a velocity of nearly 10 ft./sec., and by using a small clearance of $\frac{1}{16}''$. An example is given in Fig. 16 of a case where the punch penetrated only half way into the plate. In this figure is shown a cross section at 50 magnification; after a polish and etch, of the region adjacent to the surface of a highly localized shear strain. The plug, in lower half of picture, is firmly joined to the plate, shown in upper half of picture. The white band of metal joining the punch and plate is martensite (tetragonal lattice) which is characteristic of steel rapidly quenched from the high temperature face-centered phase. In the present case the rapid quench was provided by the adjacent material. The high temperature was attained by the concentration of shear strain characteristic of high speed shear deformation. Computation shows that a shear strain of 5 would raise the temperature 1000°C. From a consideration of the distance which the plug has been pushed ($\frac{1}{8}''$) and from the thickness of the white band ($1/800''$), it is estimated that the material in the white band has suffered a shear strain of nearly 100.

Calendar of Meetings

January

14–15 American Physical Society, New York, New York
14–15 American Association of Physics Teachers, New York, New York
14–15 Electron Microscope Society of America, New York, New York
25–27 Institute of Aeronautical Sciences, New York, New York
31–February 2 American Society of Heating and Ventilating Engineers, New York, New York

February

14–17 Technical Association of the Pulp and Paper Industry, New York, New York
18–19 American Philosophical Society, Philadelphia, Pennsylvania
25–26 American Physical Society, Metropolitan Section, Brooklyn, New York
26 American Mathematical Society, New York, New York

March

2– 4 Optical Society of America, New York, New York

April

12–15 Electrochemical Society, Milwaukee, Wisconsin
20–22 American Philosophical Society, Philadelphia, Pennsylvania
27–29 American Physical Society, Pittsburgh, Pennsylvania
28–29 American Mathematical Society, New York, New York
28–29 American Mathematical Society, Chicago, Illinois
29 American Mathematical Society, Berkeley, California

May

12–13 Acoustical Society of America, New York, New York

The Problem of Reduction of Vibrations by Use of Materials of High Damping Capacity

ANDREW GEMANT
The Detroit Edison Company, Detroit 26, Michigan
(Received September 13, 1943)

This paper deals with considerations concerning the reduction in amplitude of unwanted vibrations of machinery parts through the use of materials of high damping capacity. The problem is to reconcile the requirement for high damping with that for high mechanical quality (strength, fatigue resistance, etc.). Two ways are suggested: (1) the use of a material whose decrement generally is low but rises rapidly as the stress increases; (2) the use of a material whose damping capacity is low but rises to high peaks in certain frequency ranges. It is shown by numerical computation in two instances, namely, turbine blade vibrations and crankshaft oscillations in engines, how the suggested methods would work out in practice.

1. THE PROBLEM

IT has been recognized by several investigators in recent years that one of the means of reducing unwanted resonance vibrations of machinery parts is to produce the parts from material having an internal friction high enough to keep the amplitude of vibrations at a required low level. This statement in connection with vibration of turbine blades was brought out clearly in a paper by Allen[1] and one by Kroon.[2]

The major requirement of the problem is to a certain extent paradoxical. It is probably true to say that the more a solid material approaches the ideal state, for instance, the perfect crystal lattice, the higher its quality, measured in terms of its tensile strength, impact strength, fatigue, heat resistance, and so on. It is known, for instance, that a perfect lattice would have a tensile strength far exceeding experimental values; this discrepancy can be explained only by imperfections of real crystals. Simultaneously, however, a material that approaches perfect homogeneity will have a low internal friction. Thus, a perfect lattice would have a very high plastic resistance, actual plastic resistances being explainable only by imperfections. Hence, generally speaking, the requirement of high internal losses will reduce the level of quality of the material.

Thus the question arises, how to solve this dilemma. Its final solution will probably be in the field of practical metallurgy, being a matter of trial and error in attempting to find a middle course—that is, increasing the damping sufficiently without reducing other mechanical qualities. That this condition is attainable is shown, for instance, by certain chromium steels or manganese base alloys of relatively high damping capacity and excellent mechanical properties. However, it might be useful to look into the matter from a physical angle and discuss the possible theoretical approaches to a solution.

This paper will indicate two possible ways of specifically increasing internal friction with the minimum reduction in quality of the material generally. First, these two methods will be described briefly, and this section will be followed by some numerical illustrations for practical cases. These cases are steam-turbine blade vibrations and crankshaft vibrations in Diesel engines. The purpose of the paper is to stimulate research in the metallurgical (and plastics) fields, to be aimed at the production of materials having the described desirable damping characteristics.

It is well known[3] that the logarithmic decrement δ, which measures internal friction of any material, is not a constant, but varies with a number of factors, such as stress, frequency of vibrations, temperature, pretreatment, etc., of which the first two will be dealt with in this paper. The variation of the decrement with these factors is in some cases small, in others pronounced. In the latter cases it is possible that δ in nearly the whole range of a certain variable is relatively low, except in a more or less narrow range of the variable in question. It follows from

[1] R. C. Allen, Trans. A.S.M.E. **62**, 689 (1940).
[2] R. P. Kroon, Trans. A.S.M.E. **62**, A161 (1940).
[3] A. L. Kimball, Trans. A.S.M.E. **63**, A135 (1941).

the mechanism of losses that the greater the deviation of a material from the ideal crystal, the wider will be the range of several variables in which the decrement is high. In those cases where δ rises to high values in only limited ranges of the variables, the departure from ideal structures is the least extensive. In keeping with the statements made in the foregoing paragraphs it is thus desirable to select materials of this latter type. It is necessary only that the ranges of variables for which δ rises to high values should coincide with those ranges that actually exist under the working conditions for which an elimination of vibrations is required.

2. STRESS-SENSITIVE DECREMENT

While δ is often independent of stress in the low range, it rises to high values for high stresses. This behavior is particularly pronounced for losses due to the plastic mechanism.[4] It is obvious that such a type of behavior is particularly suited for eliminating vibrations, as it is exactly the high stresses and correspondingly high strains that characterize dangerous vibrations due to resonance. The ideal is to have a low decrement for low stresses and a sharp increase for high stresses, and the closer such a relationship is approximated, the better. We then have low loss, hence high quality material that has the ability to increase, as it were, its decrement only when danger of fatigue arises, that is, at high stresses. In consequence, high resonance amplitudes will be automatically reduced to tolerable values.

Theory shows that for plastic type losses the logarithmic decrement δ might be expected to increase with the stress somewhat as a hyperbolic cosine. For stresses not too high this might be approximated by

$$\delta = \delta_0 + \beta S^2, \tag{1}$$

with S = stress, δ_0 = constant decrement for low stresses, and β = a factor measuring the steepness of the decrement *vs.* stress relationship. In the numerical calculations below, Eq. (1) will be assumed to hold. Experimental data[3] on various alloys show that relation (1) is well approximated in a number of cases, for instance, for Cr-Mo-Va steel and for *n*-153 at 75 F.[4a] In other cases the exponent of S is by no means always 2—in fact, often less, frequently less than unity. Materials of this latter type are undesirable for our purpose.

Lazan[5] recently measured the decrement for metals and synthetic plastics. He emphasizes the higher losses of the latter group as compared with the former, also the possibility of replacing metal in certain aircraft parts by plastics in order to reduce stresses at resonance. The figures he gives for δ of Duralumin, laminated Bakelite, laminated phenolic, and Lucite show a pronounced increase of δ with stress, but according to a power much less than 2. Further efforts in the field of metallurgy and the technology of plastics will be necessary in order to increase the steepness of the δ *vs.* stress characteristic. Recent advances, such as Eyring's theory of flow processes[6] or Mark's theory of the elasticity of long chain molecules,[7] will be useful guides in approaching this aim.

Since by our assumption the first term in Eq. (1) should be small as compared with the second for high stresses, it will be sometimes sufficient for approximate computations if an equation of the type

$$\delta = \beta S^2 \tag{2}$$

is used.

An order of magnitude of 10^{-10} for β appears to be a reasonable requirement, since it would mean that the decrement would become 0.04 at the fairly high stress of 20,000 lb. per sq. in. At low stresses, say, 3000 lb. per sq. in., δ would drop to the low level of 0.001. Thus, for the material to be a low loss material, δ_0, too, should not be larger than 0.001. The use of Eq. (2) instead of (1) is then justified for stresses above, say, 5000 lb. As an example, 13 percent Cr iron approaches this behavior, δ rising from 0.002 to 0.024.[4a]

3. FREQUENCY-SENSITIVE DECREMENT

The second variable to be dealt with herein is the frequency of vibrations. In the majority of

[4] Andrew Gemant, J. App. Phys. **14**, 204 (1943).

[4a] C. Schabtach and R. O. Fehr, "Damping of engineering materials," Annual Meeting of A.S.M.E. (New York, Dec. 1943).

[5] B. J. Lazan, Modern Plastics **20**, No. 3, 83 (1942).

[6] S. Glasstone, K. J. Laidler, and H. Eyring, *Theory of Rate Processes* (McGraw-Hill, New York, 1941), p. 477.

[7] H. Mark, Am. Scientist **31**, 97 (1943).

materials investigated, the variation of δ with frequency is only slight, but quite a number of cases have been found in which δ rises to a more or less sharp maximum in a definite frequency range. Both metals (brass) and non-metals (glasses) exhibit this behavior.

In engineering practice dangerous vibrations occur in a certain frequency range in which exciting frequencies and natural frequencies of certain parts happen to coincide. It is often possible by suitably selecting the shape of parts to avoid the occurrence of such coincidence. However, this is not always possible or practical. In such cases it might be helpful to use a material that exhibits a decrement peak in exactly the same frequency range in which dangerous resonance vibrations might take place. The material should generally exhibit a decrement as low as possible, thus approaching ideal conditions, but should "automatically" regulate the amplitude in the dangerous frequency range.

The mechanism of frequency-sensitive losses is usually of a thermal nature. There are several possible mechanisms by which thermal damping can occur; the theory of these has recently been developed by Zener.[8] One mechanism takes place as a result of the variation of compression (and extension) from point to point in the specimen. Experiments show that losses of this kind can be fairly high (decrements of the order 0.01). However, it is doubtful whether they can be utilized for the present purpose as the frequency peak of δ for transverse vibrations of bars is located at rather low frequencies.

Ferromagnetism is a second possible source of frequency-dependent losses.

A third source is the presence of separate grains (crystallites) in the material. While the experimentally observed decrements due to this source are rather low, an advantage of this mechanism is the variability at will of the frequency at which δ has a maximum. Zener finds for this frequency ν_0

$$\nu_0 = 3\pi\kappa/2\rho c a^2, \qquad (3)$$

where κ = heat conductivity, ρ = density, c = specific heat, and a = average linear dimension of grains. The product $\kappa/\rho c$, called thermal diffusion constant, varies for metals in the range from 0.05 to 2 abs. units. The chief variation, and one that may be effected intentionally, is that of a, the grain size. Hence, the grain size for the material of certain parts may be selected in such a manner as to yield maximum decrements at critical resonance frequencies.

This variation of δ with the grain size should be rather sharp according to theory; i.e., the decrement should drop to low values at both sides from the maximum. This is not very clearly fulfilled in practical cases, the reason being that usually grains of various sizes are present. The more homogeneous a metal with regard to grain size—the closer it approaches physically ideal conditions—the sharper will be the maximum. The manufacture of materials having characteristics as prescribed by the requirements just mentioned will probably be aided by recent advances made in powder metallurgy.[9]

4. TURBINE BLADE VIBRATIONS: CONTINUOUS ADMISSION OF STEAM

In the following sections of this paper some practical applications of the principles outlined above will be presented. The first of these refers to steam turbine blades.

The trouble experienced with turbine blades has been treated severally in the literature.[2,10] While no details can be given here, it should be mentioned briefly that vibrations of the blades in one of their natural frequencies can be excited by a pulsating steam pressure. These vibrations may give rise to high alternating stresses near the root of the blades, and these in turn may lead to fatigue and fractures, such as are sometimes encountered in operation.

The nature of vibrations may be either of two kinds. One, which is encountered with continuous admission of steam, is a typical forced vibration of constant amplitude. The second, which is encountered with so-called partial admission of steam, is an intermittent, decaying vibration. Both types will be considered.

Vibrations of the *forced type* are dealt with, as has been done by Kimball,[11] by equating input

[8] C. Zener, Proc. Phys. Soc. **52**, 152 (1940).

[9] J. Wulff, *Powder Metallurgy* (Am. Soc. Metals, Cleveland, 1942).

[10] I. Malkin, J. Frank. Inst. **234**, 355 (1942).

[11] A. L. Kimball, *Vibration Prevention in Engineering* (John Wiley & Sons, Inc., New York, 1932).

and dissipation energies of the vibrating parts. The resonance amplitude, hence the maximum stress, is then obtained in terms of the logarithmic decrement of the material.

First, expressions concerning the exciting force have to be set up. The force F due to the pulsating steam pressure, acting on a blade, is superimposed upon a stationary pressure and varies between zero and a maximum F_0. It then has the form:

$$F=(F_0/2)(1+\cos \omega t), \tag{4}$$

where $\omega=2\pi$ times frequency ν, $t=$time.

A pulsating force is generally caused by the presence of the separate nozzles. As a blade passes from one nozzle to the next, it is subjected to pressures of varying magnitude, there being a slight decrease between nozzles. The frequency of this pulsation is several thousands per second; this pressure variation will play a role when the natural frequency of the blades is of the same high order.

A fluctuation of pressure is also caused by a variation of the pressure along the circumference of the wheel. One likely cause of such variation is an uneven wear of packing between running parts, hence a variation of steam leakage along the circumference. Such variation will have a fundamental, given by the number of revolutions per second, and harmonics, the low orders of which will have reasonably high amplitudes. Since the fundamental has a frequency of the order of 30 to 60, natural frequencies of the blades having an order of from one hundred to several hundreds may well be excited by pressure variations of this nature.

It is difficult to assess the magnitude of the pressure amplitude F_0. Since the total pressure differential across each stage is of the order of 20 or 40 lb. per sq. in., the variable force F_0 will be a fraction of this, hence probably of the order of from 1 to 5 pounds.

Next, the type of vibrations performed by the blades has to be considered; this will depend upon the support of the blades. While they are rigidly fixed to the drum at their root, the tips are usually supported by shrouds. Hence a blade will be considered as a beam built in at one end and simply supported at the other. Such a beam

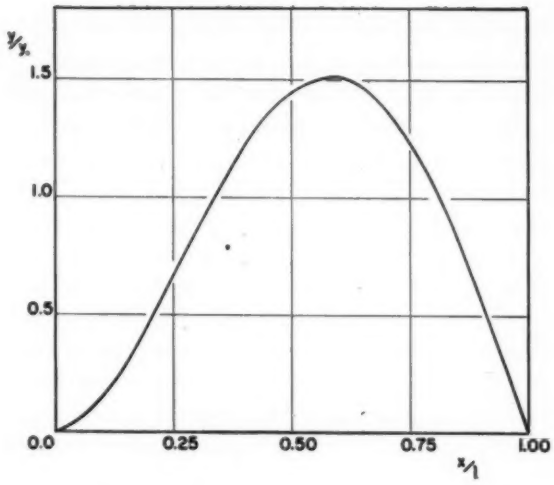


FIG. 1. Deflection (scale exaggerated) of a built-in-supported bar, the latter approximating a turbine blade.

will perform flexural vibrations in two perpendicular directions. Its displacement is in a direction either tangential to the wheel circumference, or parallel to the axis of the wheel: tangential and axial vibrations. Both may occur in actual cases. Both types may exhibit fundamental and harmonic modes; only the fundamentals will be considered below. The cross section of the blade, although often varying along its length, will be taken as constant for sake of simplicity.

If the direction along the blade is denoted as x with $x=0$ at the root and $x=l$ at the tip, and if the displacement normal to x is denoted by y, then the deflection corresponding to the fundamental will have the form:

$$y=y_0(\cos kx-\cosh kx-\sin kx+\sinh kx), \tag{5}$$

where y_0 is a constant, measuring the amplitude, and k (for the fundamental) $=3.93/l$. The variable deflection at any point is then given by $y\cos \omega t$. The deflection curve y *vs.* x is shown in Fig. 1 on an extremely magnified scale; the abscissae are x/l, the ordinates y/y_0. Of particular interest is the curvature d^2y/dx^2, since it is directly proportional to the stress at any point. The latter is given by

$$S=Ebd^2y/dx^2, \tag{6}$$

where $E=$Young's modulus and $b=$distance of the point in question from the neutral axis of the cross section. It may be seen that S has an

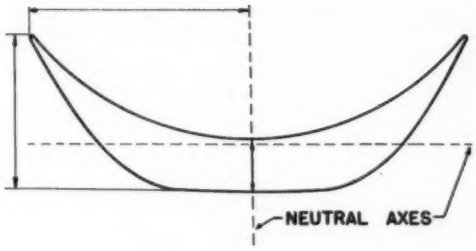


FIG. 2. Cross section of a large turbine blade, showing three main dimensions and the two neutral axes.

absolute maximum at the root, and another at about $x/l=0.6$.

The frequency of the vibration is given by

$$\omega^2=k^4EI/\rho, \tag{7}$$

where ρ = mass of bar per unit length and I = moment of area of cross section. I can be determined if the cross section is known. In first approximation, a knowledge of the three dimensions indicated on Fig. 2 is sufficient.

We now have all the data necessary for calculating the maximum stress as a function of the decrement. The maximum stress S_m is given by Eq. (6) if b is replaced by h, the maximum distance from the neutral axis, and d^2y/dx^2 by its value for $x=0$, namely, $2k^2y_0$:

$$S_m=2Ehk^2y_0. \tag{8}$$

It should be recalled that for decrements dependent on stress Eq. (1) was assumed to hold; Fig. 3 shows δ as a function of the stress.

The input energy per unit length and per cycle is

$$(\pi F_0/2l)\int_0^l ydx. \tag{9}$$

The dissipation can be shown to be 2δ times the total vibrational energy, hence is given by

$$\rho\delta\omega^2\int_0^l y^2dx. \tag{10}$$

In the last equation δ is replaced by Eqs. (2) and (6), yielding

$$(\rho\beta\omega^2E^2h^2/2)\int_0^l\left(\frac{d^2y}{dx^2}\right)^2y^2dx, \tag{11}$$

where the average of b^2 over the cross section is approximated by $h^2/2$. In equating Eqs. (9) and (11), the ratio of the two definite integrals is needed. For the case under consideration this ratio will be

$$\frac{\int_0^l\left(\frac{d^2y}{dx^2}\right)^2y^2dx}{\int_0^l ydx}=1.37k^4y_0^3. \tag{12}$$

Hence, equating (9) and (11) yields the following result:

$$S_m^3=1.2hlF_0/\beta I. \tag{13}$$

This result, as already pointed out, involves several approximations, but suffices for the question at hand.

Equation (13) contains the shape factor hl/I, constant for a given blade. It can thus be seen that the maximum stress is directly proportional to the third root of the pulsating force, and inversely proportional to the third root of the factor measuring damping. The equation gives the stress in lb. per sq. in. if h and l are measured in in., F_0 in lb., and I in in.^{4} It holds for the case where δ depends on stress, as postulated in Section 2.

In case the decrement is independent of stress (such a decrement will be denoted by $\bar{\delta}$ in the following), a relation other than (1) is valid, namely:

$$S_m=0.18hlF_0/\bar{\delta}I. \tag{14}$$

This can be deduced in the same manner as (13). The decrement $\bar{\delta}$ may stand either for one that is essentially independent of frequency or one that shows a frequency peak. According to the postulate in Section 3, the latter type appears advantageous; the frequency at which $\bar{\delta}$ is a maximum should coincide with the value of ν given by Eq. (7).

Numerical examples follow, referring to two particular sizes of blade. No. 1 represents a medium sized blade while No. 2 represents a large blade used only in the low pressure stage of large turbines; the latter type of blade occasionally causes trouble. The data referring to these two types are given in Table I.

TABLE I. Characteristic data on two sizes of blade.

	Length, in.	Moment of area, in.*	Unit mass, lb./in.	Max. distance from neutral axis, in.
No. 1 (medium)	6	0.005 (tangential)	0.035	0.3 (tangential)
No. 2 (large)	30	0.01 (tangential)	0.15	0.5 (tangential)
		0.15 (axial)		1 (axial)

* See reference 4.

Table II gives the maximum stresses from Eq. (13) in which 10^{-10} was used for β, and the values 1 and 5 lb. were assumed for F_0, these figures being of the correct order, as pointed out earlier in this section. It can be seen that, although the material has a negligibly small decrement at low and even medium stresses, owing to its peculiar $\delta - S$ characteristic, it keeps the maximum stress at a tolerable magnitude (except the figure 45,000 which is too high).

Let us now see what magnitude a stress-independent decrement $\bar{\delta}$ must assume in order to yield the same maximum stresses as shown in Table II. These values can be computed by combining Eqs. (13) and (14). The results are shown in Table III. It can be seen that the

TABLE II. Maximum stress in two blades specified in Table I.

	Tangential		Axial	
	$F_0=1$	$F_0=5$	$F_0=1$	$F_0=5$
No. 1 (medium)	16,000	28,000	—	—
No. 2 (large)	26,000	45,000	15,000	26,000

TABLE III. Values of $\bar{\delta}$ required for stresses specified in Table II.

	Tangential		Axial	
	$F_0=1$	$F_0=5$	$F_0=1$	$F_0=5$
No. 1 (medium)	0.004	0.012	—	—
No. 2 (large)	0.01	0.03	0.0035	0.01

figures range from 0.003 to 0.03, thus are much too large for high quality steel in case they signify over-all values. One way of having low over-all values but higher ones at operating conditions is that pointed out in Section 3, namely,

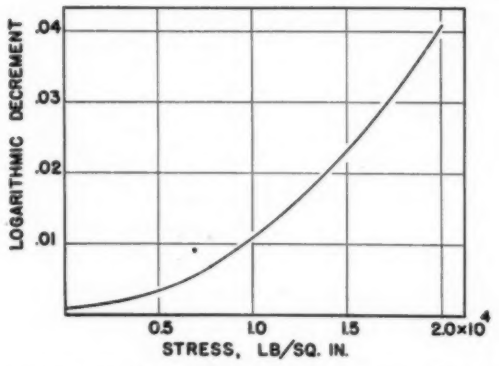


FIG. 3. Stress-sensitive logarithmic decrement after Eq. (1) with $\delta_0 = 10^{-3}$ and $\beta = 10^{-10}$.

if $\bar{\delta}$ is made to have a peak near the resonance frequency.

Table IV shows the resonance frequencies of

TABLE IV. Resonance frequencies, c.p.s., for blades specified in Table I.

	Tangential	Axial
No. 1 (medium)	2700	—
No. 2 (large)	75	290

the two blades in question, as computed from Eq. (7) (E for steel $= 29 \times 10^6$ lb. per sq. in.).

As mentioned in Section 3, $\bar{\delta}$ in polycrystalline metals will reach a maximum at a frequency ν_0, given by Eq. (3). This equation allows the calculation of a, the grain dimension, for the frequencies given in Table IV. The figures for this dimension, computed for steel, vary from 0.2 to 1 mm, hence are in a range that is within practical possibilities.

Experimental values of $\bar{\delta}$, obtained on polycrystalline metals, usually are not higher than 0.001; this would be insufficient for the present purpose, as can be seen from Table III. Theory shows that the effect increases with increasing difference of the elastic modulus as measured in different crystallographic direction. There is, therefore, a practically possible way by which the frequency effect on $\bar{\delta}$ can be made larger, and subsequently utilized for damping out unwanted vibrations at specific frequencies.

5. TURBINE BLADE VIBRATIONS: PARTIAL ADMISSION OF STEAM

The second type (intermittent, decaying type) of blade vibration is encountered in cases of

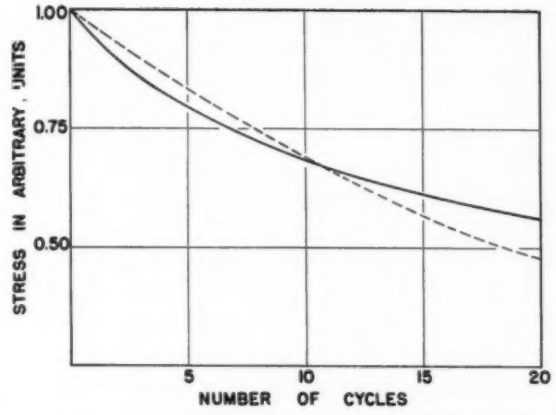


FIG. 4. Decay of free vibrations after Eq. (20) (full line); exponential (dotted line) shown for comparison.

partial admission, as analyzed by Kroon and others.[1,2] In addition to the continuous force acting on the blade during admission, there are two impulses, namely, that which occurs when the admission starts and that which occurs when it stops. Vibrations may be built up if the two impulses act during correct phases of the motion, thus performing work on the blades. The vibrations will decay during both the loaded and unloaded parts of the revolution. Equilibrium will be reached when input and dissipation are equal. In this way the maximum stress can be evaluated in terms of the decrement. The method outlined by Kroon will be used below for a decrement depending on stress according to Eq. (1).

Vibrations of this type build up easily, since the resonant frequencies lie very close to each other. The condition of resonance is that the natural frequency of the blade ν_0 should be an integer multiple of the number of revolutions per second:

$$\nu_0 = nr/60, \tag{15}$$

where n = any integer, and r = r.p.m. of drums. In this case the percentage difference between the two nearest resonant r.p.m.'s will be $100/n$. For instance, let us take $\nu_0 = 1000$, and $r = 1200$. This is a resonant condition with $n = 50$, this being the number of cycles of blade vibrations per revolution. The percentage difference of neighboring r.p.m.'s causing resonance is then 2 (1176, 1200, 1224). It is then difficult to avoid resonance, and damping by any available method deserves consideration.

As a first step for calculating maximum stresses during partial admission vibrations, the decay with time of a free vibration will be computed. Equation (1) is assumed to hold. Denoting two consecutive amplitudes of stress, one cycle apart, by S_2 and S_1, we have

$$S_2 = S_1 \exp(-\delta_0 - \beta S_1^2) \tag{16}$$

or, in first approximation:

$$S_2 = S_1(1 - \delta_0 - \beta S_1^2),$$

which is equivalent to the differential form (n = number of cycles):

$$\frac{dS}{dn} = -\delta_0 S - \beta S^3. \tag{17}$$

In integral form we have:

$$\int \frac{dS}{\delta_0 S + \beta S^3} = -n + C, \tag{18}$$

where C = integration constant. Integrating the left-hand side, we obtain

$$\frac{1}{2\delta_0} \log \frac{S^2}{\delta_0 + \beta S^2} = -n + C. \tag{19}$$

The integration constant is now evaluated from the boundary condition $S = S_0$ for $n = 0$. After some transformations the final result is obtained in the following form:

$$S = \left(\frac{\delta_0 z}{1 - \beta z}\right)^{\frac{1}{2}}, \tag{20}$$

where

$$z = \frac{S_0^2(1 - 2\delta_0 n)}{\delta_0 + \beta S_0^2}. \tag{21}$$

As an illustration, Fig. 4 (full line) shows S as a function of n, S_0 taken as unity, $\delta_0 = 0.001$, and $\beta = 0.05$. The difference between this latter value and that of 10^{-10}, as previously used, is determined by the difference in the unit used for S. The dotted line is an exponential, drawn for comparison through the ordinate corresponding to $n = 10$. It can be seen that the curve Eq. (20) is steeper at first than the exponential but later becomes flatter.

We now utilize Eq. (20) in computing maximum stresses at partial admission. Kroon's scheme of setting up four simultaneous equations will be followed. Let us denote by S_1 the maximum stress immediately following steam admission. At the end of the admission period the stress S_2 will be less, and can be calculated from our Eq. (20):

$$S_2=\left(\frac{\delta_0 z_1}{1-\beta z_1}\right)^{\frac{1}{2}}, \tag{22}$$

where z_1 is given by Eq. (21) if S_0 is replaced by S_1 and n by the number of cycles during admission.

It is now assumed that the impulse, acting while the admission suddenly stops, causes an increase of stress by an amount equal to the static stress, corresponding to the steady steam pressure. If this amount is denoted by S_s and the resultant stress by S_3, we have

$$S_3=S_2+S_s. \tag{23}$$

The stress S_3 is now falling off according to Eq. (20), and will reach the value S_4 just before admission:

$$S_4=\left(\frac{\delta_0 z_3}{1-\beta z_3}\right)^{\frac{1}{2}}. \tag{24}$$

Here again z_3 is given by Eq. (21) if S_0 is replaced by S_3, and n by the number of cycles during the intermission period. Finally, the impulse at admission adds the amount S_s to S_4 and by this means brings it back to the original value S_1. This last statement follows from the fact that the vibration is stationary with regard to one full revolution. Hence:

$$S_1=S_4+S_s. \tag{25}$$

Equations (22) to (25) contain four unknowns: S_1, S_2, S_3, S_4. A stepwise numerical calculation in any given case can be carried out easily.

Thus, let us take $S_s=4000$ lb. per sq. in., the number of cycles during steam admission as 10, and those during intermission as 30. With the numerical values for δ_0 and β used in Section 4, the values as shown in Table V for the maximum stress are obtained. The highest stress is 15,500, thus 3.9 times the static value. The average of the maximum stress is 12,500, thus 3.1 times the static value. The conclusion may be reached that a decrement dependent upon stress is sufficiently effective in keeping the maximum stress at reasonable magnitudes.

TABLE V. Maximum stresses (lb./sq. in.) at partial admission, calculated from Eq. (20).

S_1	13,600
S_2	11,500
S_3	15,500
S_4	9,600

The other alternative, decrement dependent upon frequency, might be helpful in this case, too. An effective constant value of the decrement at the resonance frequency should be of the order 0.01, and the frequencies in question are usually of the order of 1000 or somewhat greater. As pointed out previously, the full use of both methods suggested will require special research in metallurgical fields. The question as to which of the two alternatives should be used depends chiefly upon which technical problem will be found easier to solve.

6. CRANKSHAFT VIBRATIONS IN DIESEL ENGINES

The next case to be considered is the torsional vibration of crankshafts in Diesel engines.[12,13] It is true that the use of torsion dampers on crankshafts is common practice today, but this does not preclude the interest in this case from our standpoint. Apart from the use of Diesel engines for ship drives or electric generators, they find an increasing application in locomotives; hence they are being used in widely differing fields.

The chief cause of torsional oscillations of the shaft is the irregularity of the gas pressure torque of the cylinders. Apart from the average steady torque, periodic peaks are produced that can be resolved after Fourier in harmonics, expressed in terms of the fundamental rotational frequency of the engine. If any of these harmonics coincide with any of the natural torsional frequencies of the crankshaft, forced vibrations originate that, if large amplitudes occur, might lead to fracture. Those speeds of the engine at which resonance occurs are called critical speeds.

[12] Andrew Gemant, J. App. Phys. **14**, 258 (1943).
[13] Ralph Poole, Proc. Inst. Mech. Eng. **146**, 167 (1942).

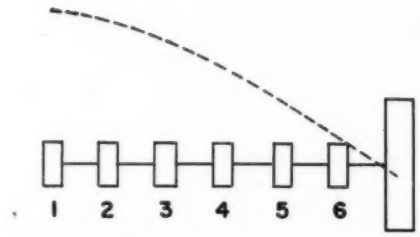


FIG. 5. Fundamental vibration of a six-cylinder Diesel electric generator; dotted line shows amplitude.

Because the several cylinders are not in phase, the alternating torques are not in phase either and must be added vectorially if the resultant is required. In most cases the resultant will be rather small as compared with the single terms. For the third, sixth, etc., harmonics (in a six-cylinder engine), however, all the torques will be in phase, so that the resultant is fairly large. Speeds of the engine at which these particular harmonics lead to resonance are the most dangerous; they are called major critical speeds. In the following example we shall consider such a major critical speed.

The scheme of the computation is in principle the same as that used in the previous sections; work input and dissipation are computed and equated; this gives a relation between decrement and maximum stress.

In order to fix our ideas, a specific case will be considered. Let us take a six-cylinder Diesel electric generator, of about the size used in a locomotive;[13a] there are often two separate identical units operating on the same locomotive. Table VI gives the chief data for such an engine.

TABLE VI. Characteristics of a 6-cylinder Diesel engine.

Diameter of crankshaft	3½ in.
Separation of cylinders	9 in.
Elasticity of each shaft section	20 in. lb./rad.
Moment of inertia of each crank	1 lb. in. sec.²
Moment of inertia of flywheel plus generator	60 lb. in. sec.²
Total power	90 hp
Running speed	1800 r.p.m.

The first step is the determination of the natural frequency of the shaft. It is the fundamental mode of vibration with which we are concerned. This mode is shown schematically in Fig. 5. In this mode the cranks rotate in one direction, the flywheel-generator assembly in the other. The node is between the sixth crank and the flywheel. The amplitude of the cranks decreases as the node is approached. The frequency of this mode can be computed by the Holzer method;[14] it is very nearly 180 c.p.s.

If the amplitude in radians of the first crank is denoted by θ, then that of all others can be computed. Table VII gives these angles in terms of certain factors f, each amplitude being given by $f\theta$. The twist of each shaft section is also given in terms of factors f', each angle being given by $f'\theta$.

TABLE VII. Amplitudes of cranks and shaft sections in fundamental mode (for engine specified in Table VI).

Cylinder No.	f	f'
1	1.00	0.063
2	0.937	0.122
3	0.815	0.173
4	0.642	0.213
5	0.429	0.240
6	0.189	0.252

The next step is the calculation of the work spent in vibrating the shaft. If T is the torque amplitude per cylinder, then the total work per cycle W is given by:

$$W=\pi T\sum f\theta, \tag{26}$$

or

$$W=\pi T\theta\sum f. \tag{27}$$

Now, T can be estimated if one knows the steady torque T_0. The latter is given by the equation:

$$T_0=5260\frac{\text{hp}}{\text{r.p.m.}}\text{ ft. lb.} \tag{28}$$

In our case (see Table VI) the value of this torque per cylinder is 44 ft.-lb. T can be estimated from T_0 by means of certain graphs.[14] Several authors, F. M. Lewis[15] and E. S. Taylor,[16] among others, have established the magnitude of torque harmonics in fractions of the average

[13a] The author is indebted to Cummins Engine Company, Columbus, Indiana, for pertinent information on an engine of this type.

[14] J. P. DenHartog, *Mechanical Vibrations* (McGraw-Hill, New York, 1940).

[15] F. M. Lewis, Trans. Soc. Nav. Arch. Marine Eng. **33**, 109 (1925).

[16] G. P. Bentley and E. S. Taylor, J. Aero. Sci. **6**, 1 (1938).

torque T_0. By using these factors, T can be found if T_0 is known.

Since the r.p.m. of our engine is 1800, the c.p.s. is 30, and the sixth harmonic of the gas pressure torque will be in resonance with the natural frequency of 180. A representative factor for the sixth torque harmonic is 20 percent.[14] Hence $T=8.8$ ft.-lb.

$\sum f$ from Table VII is 4.0, and Eq. (27) then yields $W=110\theta$ ft.-lb./cycle.

In case the frequency (and the horse power) were doubled, the third harmonic would be in resonance with the fundamental mode. In this case the graphs mentioned above show a harmonic torque of about 120 percent of the steady torque, giving $T=53$ ft.-lb. Strictly speaking, a torque component resulting from the inertia of the crank should be properly compounded with the value as given above. For the present purpose this may be disregarded, however. $W=660\theta$ ft.-lb./cycle is then obtained for the work input.

In order to compute the energy dissipation, let us first consider a single cylindrical shaft section of length l, radius R, and torsional modulus N. If the total torsional amplitude is α, then the torsional amplitude per unit length is α/l and the shear at distance r from the axis $\alpha r/l$. The shearing stress then is $N\alpha r/l$, and the shearing torque $N\alpha r^2/l$. The energy per unit area of cross section is $N\alpha^2r^2/2l$, and the corresponding dissipation $\delta N\alpha^2r^2/l$. The total dissipation is then

$$\int_0^R \frac{2\pi\delta N\alpha^2r^3}{l}dr. \tag{29}$$

Considering losses on the plastic basis (Section 2), δ is replaced by its value from Eq. (1), where

$$S=N\alpha r/l. \tag{30}$$

We then have for (29):

$$\frac{2\pi N\alpha^2}{l}\left(\delta_0\int_0^R r^3dr+\frac{\beta N^2\alpha^2}{l^2}\int_0^R r^5dr\right) \tag{31}$$

or, integrated:

$$\frac{\pi NR^4\alpha^2}{2l}\left(\delta_0+\frac{2\beta N^2R^2\alpha^2}{3l^2}\right). \tag{32}$$

We now have to sum the dissipation for the six shaft sections by replacing α in Eq. (32) by $f'\theta$. The total dissipation is then:

$$\frac{\pi NR^4\theta^2}{2l}\left(\delta_0\sum f'^2+\frac{2\beta N^2R^2\theta^2}{3l^2}\sum f'^4\right). \tag{33}$$

From Table VII, where the values of f' are listed, one obtains $\sum f'^2=0.21$ and $\sum f'^4=0.0105$.

In the following, use is made of a relation between the angle θ and the maximum shearing stress S_m. This stress will be found at the outer fibers of the sixth shaft section, the torsional amplitude of which, as can be seen from Table VII, is the largest of all. The relation referred to is:

$$S_m=0.25NR\theta/l. \tag{34}$$

By putting numerical values (with N for steel $=12\times10^6$) into expression (33) and equating it with W, the following is obtained:

$$W/\theta=7.2\delta_0S_m+3.8\beta S_m^3. \tag{35}$$

In the previous sections the maximum stress was evaluated from an assumed value of β. The relation, however, can be used as well for calculating the value of β that is necessary in order to reduce the maximum stress to an arbitrary value. Let this be 10,000 lb./sq. in., and, further, let $\delta_0=0.5\times10^{-3}$.

With $W/\theta=110$, as obtained above for 1800 r.p.m., $\beta=0.2\times10^{-10}$ results. This is five times less than the value assumed in Sections 4 and 5, the stresses produced being less severe than in those cases. A material fulfilling this requirement of relatively low β should be easy to find.

The stresses are higher, however, if—as considered above—a speed of 3600 r.p.m. is taken. W/θ then becomes 660, and with $\delta_0=0.5\times10^{-3}$, $\beta=1.6\times10^{-10}$ results. This is even higher than the value 1×10^{-10}, taken in the previous section, partly because of the fact that the maximum stress is now limited to 10,000, whereas those obtained in the previous sections were higher on the average. In any case, it can be seen that stress-sensitive decrements will sufficiently limit the stresses in shafts undergoing torsional vibrations.

The alternative mechanism—frequency-sensitive decrements—is not feasible here, as damping due to the thermal mechanism occurs only with flexural or longitudinal vibrations, and not with torsional ones.

I am much indebted to my colleague Sergius Vesselowsky for helpful discussions of the subject.

The Technique of Microradiography

S. E. MADDIGAN
Research Physicist, Chase Brass and Copper Company, Waterbury, Connecticut
(Received October 4, 1943)

Investigations have been made of the various factors considered to influence the sensitivity of the microradiographic method for the examination of alloys. Both target element and applied voltage influenced the results. Radiographs are shown to demonstrate the effect; an explanation is given in terms of the x-ray emission curves and the absorption differential curves for the phases of the radiographed alloy. Target-specimen distance was an important factor for uncollimated rays, particularly for distances of less than 6 in. with minute inhomogeneities or with low values for the absorption differential. The size of inhomogeneity, lower limit of absorption differential, and upper limit of magnification were also studied. For investigation of the distribution of elements in inhomogeneities, it was concluded that, contrary to previous reports, both target element and applied voltage must be carefully chosen after a study of the absorption differential curves. The optimum voltage range is usually 2.5 to 3.5 times the excitation potential.

INTRODUCTION

RECENT intense interest in the results obtained on metals with the electron microscope has relegated to a position of comparative obscurity the use of microradiography. While the former, through an entirely new order of resolution, opens fresh vistas to the physical metallurgist, the latter offers equal potentialities by opening the way to a three-dimensional study of the metal interior.

The importance of the additional information obtained by the radiographic method is readily seen by comparison of an optical photomicrograph and a microradiograph, taken in the same general region of an alloy of 80 percent copper, 10 percent tin, 10 percent lead, in the as-cast condition (Figs. 1 and 2). The optical method shows the existence of two phases of the copper-tin system plus dark particles of segregated lead. The radiograph, on the other hand, gives a remarkably clear picture of the dendritic growth of the metal crystals.

The disclosure of the internal structure of the alloy is dependent upon the laws of absorption of x-rays. As is well known, the relation between the intensities of the incident x-ray beam and that of the beam after transmission through a layer of material of thickness t is given by the equation,

$$I_t = I_0 e^{-\mu t},$$

where the symbols have their commonly accepted significance. Of chief concern for microradiography is the case of a layer of material of thickness t_1 and absorption coefficient μ_1 containing an inhomogeneity of thickness t_2 and absorption coefficient μ_2. The transmitted x-ray beams through the homogeneous matrix and through the region containing an inhomogeneity are related by the expression,

$$I_a/I_b = \exp\left[-t_2(\mu_1-\mu_2)\right],$$

where $(\mu_1-\mu_2)$ is the difference in linear absorption coefficient for the matrix and the inclusion. For a given alloy phase of density d, consisting of a homogeneous arrangement of two elements of amounts X and Y and mass absorption coefficient ρ_x and ρ_y, the linear absorption coefficient would be

$$\mu = \left[(X\rho_x + Y\rho_y)/X + Y\right]d.$$

For the alloy under consideration in Figs. 1 and 2, a study of the phase diagrams shows that lead is almost completely insoluble in the solid state and would, therefore, be thrown out as segregated clumps. The copper-tin solid solution, on the other hand, would solidify in a composition range of 97 percent copper, 3 percent tin, to about 76 percent copper, 24 percent tin. The linear absorption coefficients and the differential $(\mu_2-\mu_1)$ for cobalt radiation are shown in Table I.

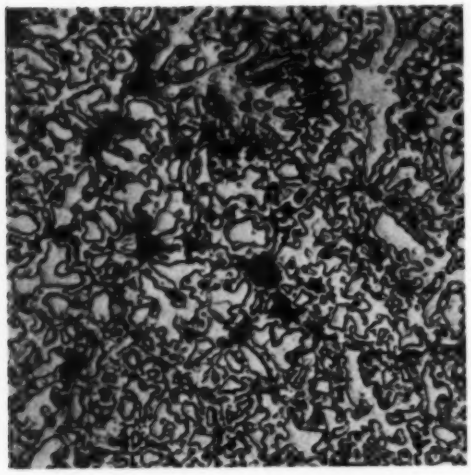


FIG. 1. Photomicrograph of 80 percent Cu, 10 percent Sn, 10 percent Pb at 60×. Pb appears as dark segregation.

TABLE I. Linear absorption coefficients.

97% Copper– 3% Tin	830
76% Copper–24% Tin	1370
Differential	540

During solidification, the crystals begin to grow with spines and branches of the copper-rich composition, but gradually the tin concentration increases until the final interdendritic material consists of the tin-rich composition. It can be seen from the foregoing that various parts of the dendritic structure will have a maximum value of $(\mu_2-\mu_1)=540$ for Co K radiation, and the relative intensities of the transmitted beams through the copper-rich and the tin-rich regions will be given by the equation,

$$I_a/I_b=\exp 540t_2.$$

Thus, because of the variation in absorption with composition, the radiograph is able to give a remarkable picture of the solidification process.

The contrast with which such variations in metal structure become visible is dependent from the above explanation only on the value of $(\mu_2-\mu_1)$ and the thickness t_2 of the inhomogeneity. The inhomogeneity may at times extend completely through the radiographed specimen, but at other times may be only a small particle buried within the matrix.

The early work on microradiography was confined chiefly to light alloys and made use of low voltage, long wave-length radiation, in order to obtain a maximum differential in absorption between the various components. In order to avoid excessive exposure times, absorption in the specimen had to be reduced by the use of extremely thin samples, and air absorption had to be eliminated by the use of vacuum equipment. Clark and his collaborators, however, demonstrated that by the use of characteristic radiations of an intermediate wave-length, such as are commonly used for x-ray diffraction work, the exposure time could be greatly reduced and the difficulty regarding the absorption differential could be overcome by a suitable choice of target material. This is indicated in Fig. 3, where the heights of the lines represent the linear absorption coefficients for the characteristic radiations of copper, molybdenum, and cobalt in a number of elements. As was pointed out by Clark and Shafer,[1] an alloy of aluminum and copper would actually give better contrast with the characteristic radiation from molybdenum rather than from copper, although from the older work, using long wave general radiation, one would expect the longer characteristic wave-length to be preferable. In the case of a copper-tin alloy,

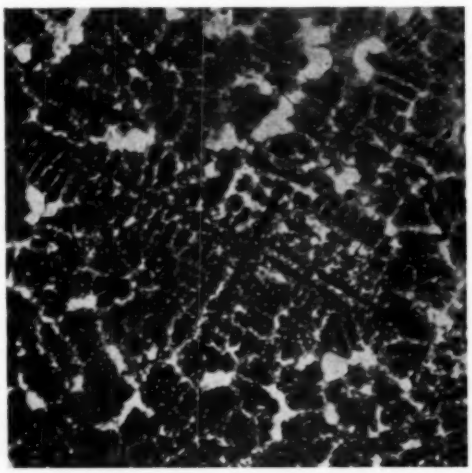


FIG. 2. Microradiograph of 80 percent Cu, 10 percent Sn, 10 percent Pb at 80×. Specimen thickness 0.002 in. Cu-rich phase appears dark; Sn-rich phase appears as white parts of dendrites; Pb appears as large white areas.

[1] G. L. Clark and W. M. Shafer, Trans. A.S.M. **29**, 732 (1941).

on the other hand, the contrast is much superior using Cu K or Co K characteristic radiation. Clark[2,3] and his collaborators have demonstrated the use of such characteristic radiations for a variety of alloys. In their discussions, however, they completely ignored the effect of the general background radiation accompanying the characteristic type.

Woods and Cetrone[4] have indicated that, for light alloys at least, radiography by means of characteristic radiations seemed no more favorable than by the use of low voltage general radiation. In their study of the optimum conditions for radiography, step tablet pictures showed increasing contrast with decreasing voltage. This was equally true for targets of tungsten, molybdenum, and copper. Their results were verified by radiographs of actual voids and metallic inclusions in aluminum alloys, and showed that for either type of inhomogeneity the contrast with tungsten at a low voltage was superior to either copper or molybdenum at a somewhat higher voltage. On the other hand, for a given voltage, the molybdenum radiation showed a slight but definite superiority in contrast over either tungsten or copper except at the lower voltages where the Mo K radiation was not excited. This work at first glance contradicts the results of Clark, but actually agrees in some respects with the latter. From the absorption values (Fig. 3) the characteristic radiation of molybdenum should give better contrast for the aluminum-copper alloy studied by Woods and Cetrone than does the characteristic radiation of copper, but as will be shown this is not true for aluminum-manganese. The explanation for the similarity between tungsten and copper radiations was not mentioned by Woods and Cetrone, but is readily understood when it is remembered that the L series of tungsten lies very close to the K lines of copper ($WL\alpha_1 = 1.473$A; $WL\beta_1 = 1.279$A; Cu $K\alpha = 1.54$A).

We are thus presented with two differing opinions, the one stressing voltage, the other

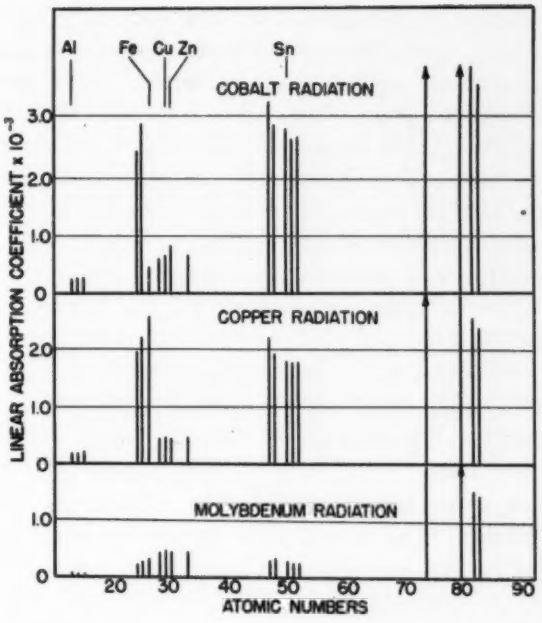


FIG. 3. Linear absorption coefficients for Co, Cu, and Mo characteristic radiations in various elements.

the target element, as the major factor. To resolve this controversial situation, it was felt desirable to carry out a further investigation on the various items which would affect the detail of the resulting microradiograph. The main items which might influence the radiographic contrast are as follows: I, voltage and target element; II, specimen-target distance; III, inclusion thickness and magnification; and IV, specimen thickness.

EXPERIMENTAL PROCEDURE

The same general experimental procedure was used as outlined by Clark.[1,2] Three x-ray tubes were available; two of the Machlett type with beryllium windows and with cobalt and molybdenum targets, and one of the older General Electric CA-4 type with Lindemann glass windows and copper target. These were operated on self-rectified voltage. The linear absorption coefficients for the three characteristic radiations in a number of elements are shown in Fig. 3.

Specimens were prepared by a procedure available in most metallographic laboratories. Flat saw-cut samples of essentially uniform gauge were mounted in Bakelite mouldings, and one surface prepared for examination with the metal-

[2] G. L. Clark and S. T. Gross, Ind. Eng. Chem. Anal. Ed. **14**, 676 (1942).
[3] S. T. Gross and G. L. Clark, Iron Age **152** (4), 44 (1943).
[4] R. C. Woods and V. C. Cetrone, Iron Age **151** (12), 52 (1943).

lographic microscope. This was necessary since in most cases it was desired to correlate the radiographs with photomicrographs. After enscribing an interesting region with a small indicating circle the moulding was split, the specimen reversed, and mounted again in a Lucite mounting. This was then placed in the chuck of a lathe and the second surface turned down parallel to the first with rather light cuts until the sample reached a gauge of about 0.010 in. From this thickness it was hand polished on emery papers of varying grades, until approximately the desired gauge was reached (usually 0.005 in. or less), then finished off on French papers to 3/0 grade, using an oil film for the final stages. In most cases this final polishing eliminated any scratches otherwise visible in the radiograph. The entire procedure required about the same time of preparation as that for a good metallographic specimen.

After removal of the thinned specimen from the Lucite, exposures were made with the selected region on the sample placed behind a $\frac{1}{8}$-in. hole in a brass plate. No collimation was considered necessary, as exposures were made off the end of a line focus of about 1-mm width. Eastman Type 548-0 spectrographic plates were used with the emulsion separated from the metal specimen only by an intervening single layer of black paper. Development was carried out at 65°F for 1.5 minutes in D-8 contrast developer. Several exposures were made on each plate simply by moving plate and containing envelope a short distance between exposures. In making comparisons between different conditions of voltage, etc., radiographs with as nearly as possible equivalent over-all densities were used. In each case this was also the exposure with the greatest contrast between phases.

The radiographs were enlarged by the use of a metallographic microscope. Enlargements were tried both by reflection from the emulsion surface using the microscope exactly as for the study of a metal specimen, and also by transmission, by placing the photographic plate on the stage of the microscope and using a 45° mirror to deflect the rays from the regular microscope light source through the radiograph. As was to be expected, radiographic exposure times necessary for the reflection method were much longer than those which gave the same amount of detail by transmission. In other ways, however, the reflection method was quite suitable. Due to the much shorter exposure time required, all of the work done in this investigation was carried out by transmission enlargement. All enlargements were made on the same type of metallographic plate, and all prints for a given specimen were made on the same grade of paper.

Radiographs were made with all three target elements with applied voltages from 17 to 42 kv. Voltages below 17 kv were found impracticable because of the long exposure time. Most of the exposures were made with a specimen-target distance of 4 in. but some tests were made with distances from 2 to 6 in.

A magnification of 100× was chosen for most of the work after the effect of higher magnifications had been investigated.

RESULTS AND DISCUSSION

I. Voltage and Target Element

In order that the results would have somewhat broader application than those of Woods and Cetrone, two types of alloy were studied: (1) copper base alloys, and (2) an aluminum base alloy.

Results have already been shown in Figs. 1 and 2 for the case of an alloy of 80 percent copper, 10 percent tin, 10 percent lead. Clark[2] investigated a similar alloy and showed that even with molybdenum radiation a slight indication of the tin-rich regions was obtained, although according to the absorption values (Fig. 3) the tin-rich phase should be highly visible either with cobalt or copper, but with molybdenum should be either invisible or slightly reversed. Results similar to those of Clark were observed in the present investigation. Clark had suggested that the discrepancy was due to a slight absorption of lead in tin. Due to the process of solidification of the alloy, as previously discussed, it was felt that this could not be the explanation, but that the true reason was probably linked up with the general background radiation which had been ignored by Clark, but which, according to Woods and Cetrone, is the predominant part of the x-ray emission for radiographic purposes. In

order to investigate the subject further, an alloy of 62 percent copper, 3 percent tin, 35 percent zinc was radiographed in the forged condition. A study of the ternary diagram[5] showed that this alloy had a solidification range of 1.3–4 percent tin with only a small variation in copper and zinc content. Since the copper and zinc mass absorptions are almost identical, except in the narrow range between the K absorption levels, the linear absorption coefficients for this alloy were calculated considering it as a copper-tin alloy and ignoring the zinc. Upon forging and cooling to room temperature, the alloy consisted of grains of the alpha-phase with intergranular eutectoid of alpha and gamma in agreement with the phase diagram. The alpha-phase was considered as 1.3 percent tin and the gamma-phase as 4 percent tin for calculation of the absorption coefficients. The specific gravity is not greatly different for the two phases, and was taken as 8.4 for both. The absorption differential $(\mu_2-\mu_1)$ is shown for various wave-lengths in Fig. 4.

Radiographs were also made of a commercial aluminum alloy (Type 3S) containing 1.25 percent manganese. At room temperature this alloy consists of two phases,[6] one phase with about 100 percent aluminum, and the second phase with a composition $MnAl_6$ (approximately 25 percent manganese). In order to calculate the absorption coefficients, the specific gravity of the first phase was taken as that of aluminum (2.70). The $MnAl_6$ phase must have a value somewhat higher than this, but since no specific gravity or crystal structure data were found in the literature a simple ratio value of 3.2 was chosen as an approximation for the second phase. The absorption differential values are shown in Fig. 4.

Radiographs were made of these alloys at several voltages ranging from 42 kv down to 17 kv with both molybdenum and cobalt radiations. The results for the copper alloy are shown in Fig. 5. Using cobalt radiation, the tin-rich phase was shown clearly at all voltages although the contrast was slightly greater for the lower voltages. The change was most noticeable above 30 kv. For molybdenum radiation at 42 kv the tin-rich phase showed only very faintly, but as the voltage was decreased the contrast* became greater until, at 20 kv, contrast with molybdenum was almost as good as with cobalt radiation, and at 17 kv no difference could be seen

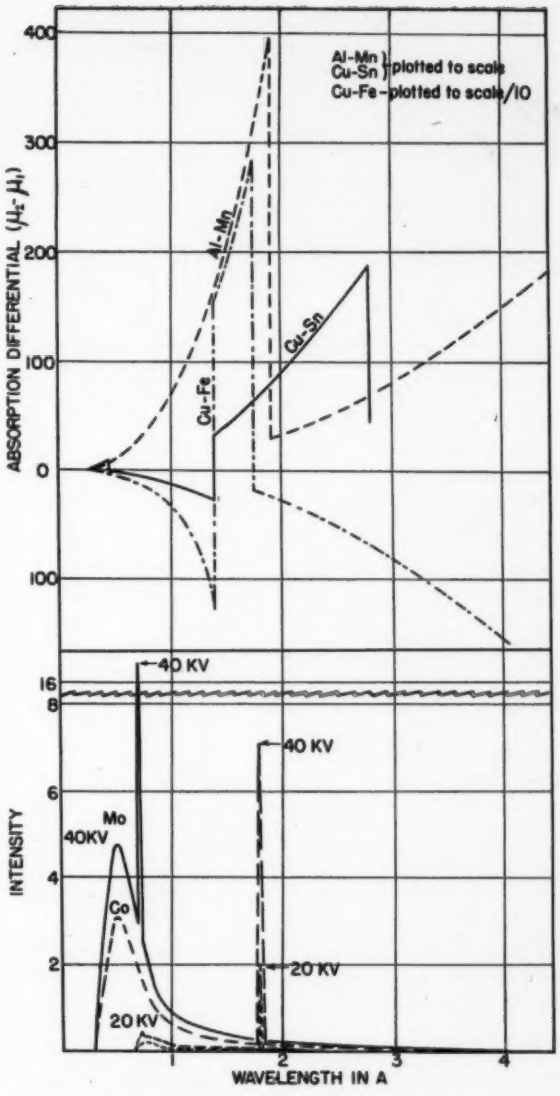
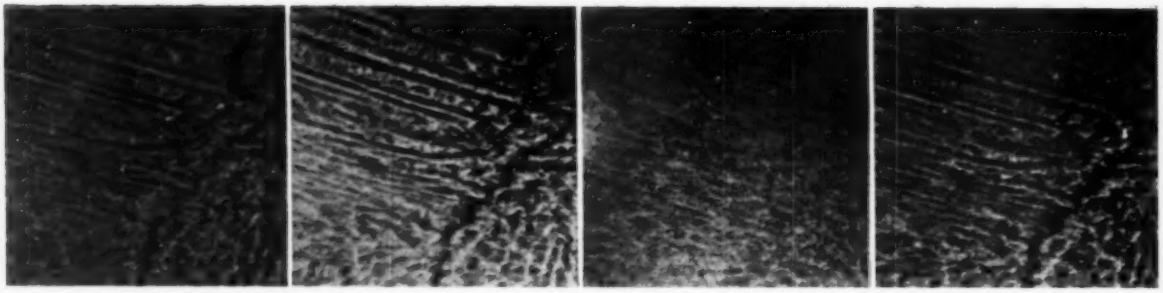


FIG. 4. X-ray emission curves for Co and Mo targets at 20 and 40 kv, and absorption differential curves for alloys of Al-Mn, Cu-Zn-Sn (Cu-Sn), Cu-Zn-Fe-Si (Cu-Fe).

[5] O. Bauer and M. Hansen, Zeits. f. Metallk. 22, 405 (1930).

[6] A.S.M. Metals Handbook (American Society for Metals, 1939).

* Some confusion may result from the relatively short range of printing paper as compared to the photographic plate. This sometimes results in as good over-all detail for prints from a plate with low contrast as from one with high contrast. The criterion for sensitivity of the radiograph, however, is contrast and this can be accurately judged only by examination of the enlargement negatives from which the displayed prints were made.



(a) (b) (c) (d)

FIG. 5. Microradiographs of Cu-Zn-Sn alloy at 50×. Cu-rich phase appears dark; Sn-rich phase appears light grey to white.
(a) Co radiation, 42 kv (b) Co radiation, 17 kv
(c) Mo radiation, 42 kv (d) Mo radiation, 17 kv

between the two types of x-rays. Similar results were observed for the two phases in the aluminum-manganese alloy (Fig. 6). In this case, even at 17 kv, results with molybdenum and cobalt were not equivalent, but as indicated in Table II the contrast produced by cobalt at 30 kv was equal to that with molybdenum at 17 kv. These results do not agree with those for aluminum alloys obtained by Woods and Cetrone.

The results are readily explained by a correlation of the absorption differential curves with the x-ray intensity distribution curves for the two target elements.† For a given voltage, the intensity distribution of the background radiation should be about the same for different targets with the total intensity varying according to the law, $E=KZV$. Transmission through a layer of material will reduce the radiation at a given wave-length by the same factor regardless of target element, and similarly the ratio between the transmissions through two different phases will remain constant regardless of target. Since the photographic contrast is influenced by the ratio between intensities, rather than by their arithmetical difference, it is readily seen that the contrast produced by integrated background radiation alone would depend only on the voltage and not upon the target element. Considering only the general background curves in Fig. 4, it can be seen that a change in target element would change only the exposure time for a given voltage, but reducing the voltage from 40 to 20 kv throws a much larger proportion of the radiation into a region of high absorption differential, and thus the ratio of the integrated intensities of transmitted beams $\int I_2 : \int I_1$ would be shifted further from unity.

When the characteristic radiation is added a marked change occurs. At 40 kv the high intensity Mo K radiation is in a region of low absorption differential for both alloys mentioned above, and, therefore, $\int I_2 : \int I_1$ is shifted toward unity. As a result, the contrast is much less than would be the case for the general radiation alone. On the other hand, the Co K radiation lies in a

TABLE II. Ratings on radiographs.

	Alloy No. 1 98.5% Al, 1.5% Mn		Alloy No. 2 62% Cu, 35% Zn, 3% Sn		Alloy No. 3 56.3% Cu, 40.5% Zn, 2% Fe, 0.2% Si, 1% Sn		
kv	Cobalt	Molybdenum	Cobalt	Molybdenum	Copper	Cobalt	Molybdenum
42	*C*	*E*	*E*	*D*	*B*	*F*	*F*
30	*B*	*D*	*B*	*D*	*A*	*F*	*E*
20	*B*	*C*	*A*	*B*	*A*	*F*	*D*
17	*A*	*B*	*A*	*A*	*B*	*E*	*C*
			Exposure time in minutes				
42	1.5	0.7	30 (6 in.)	15 (6 in.)	15	10	10
30	2	1.5	30 (6 in.)	20 (6 in.)	30	30	20
20	12	4	10 (2 in.)	15 (2 in.)	55	45	40
17	14	20	90 (2 in.)	90 (2 in.)	180	120	120

N.B. Radiographs are rated for each specimen in order of merit with respect to contrast between the phases. The decrease in contrast is not necessarily the same from one step to another, nor in the corresponding step from one specimen to another. Exposure times listed are for tube currents of 15 ma except at 42 kv, where the current was 10 ma. Target-specimen distance was 4 in. except where otherwise noted.

Relative exposure times are affected by: possible slight shifts of specimen perpendicular to x-ray beam; slight variations in developing technique; change from beryllium to Lindemann glass window in x-ray tube.

† The general radiation curves were deduced from the curves for a tungsten target in Compton and Allison's *X-Rays in Theory and Experiment* (D. Van Nostrand Co., Inc., 1935), page 38, using the equation x-ray efficiency $=1.1\times10^{-9}ZV$. The intensities of the characteristic radiations were obtained from the *International Critical Tables* 4, pp. 33–34.

region of high absorption differential, and thus for a cobalt target $\int I_2 : \int I_1$ is shifted further from unity, with a resultant increase in contrast.

At 20 kv the Mo K intensity is negligible, while the Co K is much less intense than before but is still high compared to the background, with a resultant shift in the transmitted intensity ratio. The contrast for a molybdenum target is that provided by the general radiation, while for cobalt it is somewhat greater.

At still lower voltages the effect of the cobalt characteristic radiation becomes smaller so that the results for molybdenum and cobalt are almost identical.

The behavior with aluminum-manganese and copper-tin alloys should be in general the same but with some variation between them since the wave-length region of high absorption differential is somewhat different in the two alloys. Reference to Table II and comparison of Figs. 5 and 6 shows clearly that this is the case.

The case of an aluminum-copper alloy studied by Clark and Shafer is quite different in that the Mo K radiation now lies in a region of high absorption differential, while the corresponding value for the Co K rays is low. Thus, the greater contrast would be obtained with the molybdenum target and at the higher voltages.

The above argument is demonstrated more clearly by a consideration of the actual intensity for the transmission through the two phases of the aluminum-manganese alloy. Figure 7 shows how the contrast would be decreased by the Mo K rays at 40 kv since the transmission ratio of the high intensity characteristic rays is 11 : 14.4, which is higher than the ratio of the integrated backgrounds. The cobalt characteristic rays have a transmission ratio 0.004 : 1.1 which decreases the ratio of $\int I_2 : \int I_1$, and thus increases contrast. In Fig. 8 for 20 kv the ratio of the background areas is seen to be much more favorable for contrast than was the case at 40 kv. The

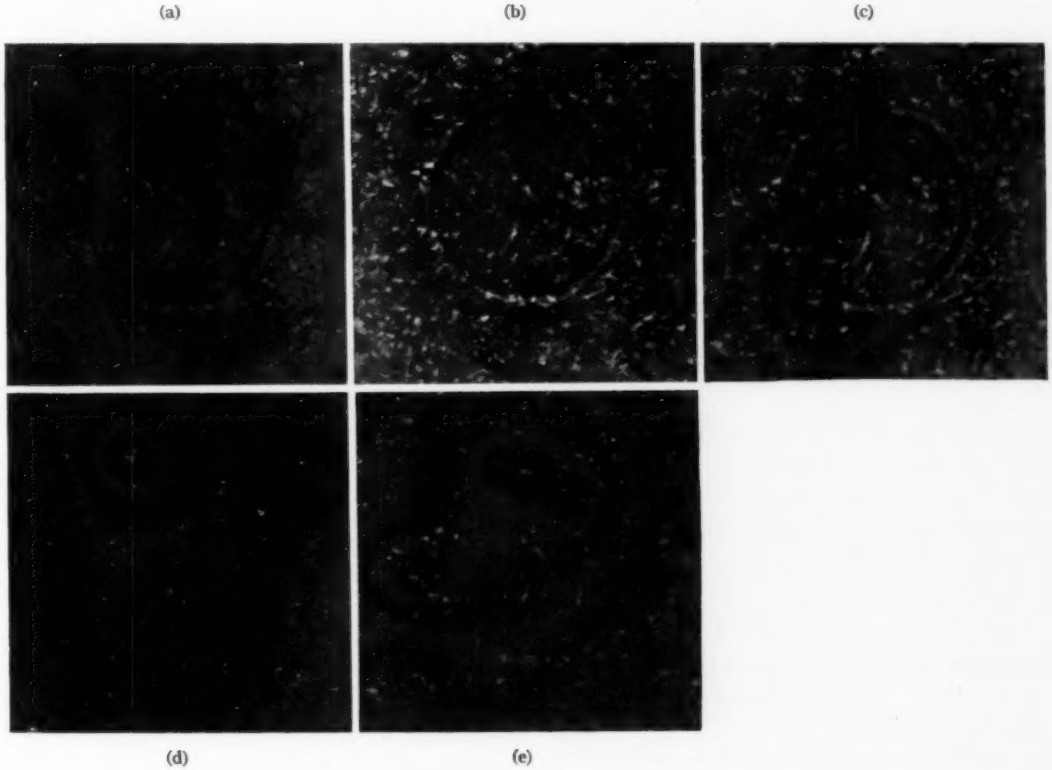


FIG. 6. Microradiographs of Al-Mn alloy at 60×. Al-rich phase appears dark; Mn-rich phase appears light grey to white.
(a) Co radiation, 42 kv (b) Co radiation, 30 kv (c) Co radiation, 17 kv
(d) Mo radiation, 42 kv (e) Mo radiation, 17 kv

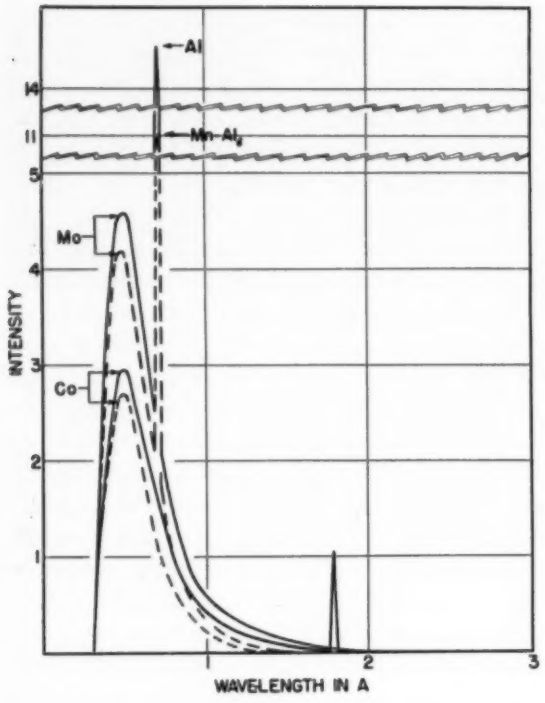


FIG. 7. Transmitted intensity curves for the two phases of the Al-Mn alloy for Mo and Co targets at 40 kv (particle thickness assumed = 0.01 cm).

Mo K intensity is now negligible while the ratio of the transmitted Co K rays is 0.0012 : 0.295, which adds considerably to the total contrast for the latter. Similar evidence is presented if the transmitted intensity curves for the copper-tin alloy are studied. The longer wave-lengths are more effective than shown above since due to the self-rectified operation of the x-ray source, the fluctuating voltage shifts the intensity distribution toward the long wave end of the curve.

To complete the study it was necessary to test the effect of a characteristic radiation on the long wave side of the absorption differential discontinuity such as occurs for aluminum-manganese at 1.89A or for copper-tin at 2.77A. A brass containing iron was studied with such a jump for the copper-zinc *vs.* iron at 1.739A. The Cu K radiation on the one side of this has a high differential, while Co K on the other side of the jump has a low value.

As shown in Fig. 4, the absorption differential curve for this alloy has an extremely high, narrow band with values running from about 1500 at 1.377A to 2900 at 1.739A. Throughout the rest of the x-ray spectrum the absorption differential is negative and has much smaller absolute values.

Radiographs were made as before with molybdenum, copper, and cobalt radiations, with applied voltages varying from 42 to 17 kv. For this alloy, the Cu K radiation is in the region of extremely high absorption differential, while the Mo K has a rather small negative value, and the Co K also has a negative but numerically larger one. Figure 9 shows the results for these tests. At all voltages, radiographs with the copper target showed precipitated iron or iron-rich particles with an exceptional sharpness and high contrast. With molybdenum radiation, no indication of these particles was found at 42 kv, since the negative absorption differential for the high intensity Mo K rays completely counteracts any slight effect which might be produced by the background radiation in the range 1.377–1.739A. At lower voltages, the Mo K radiation becomes weaker so that at 30 kv the iron-rich particles began to be visible, and at 17 kv very good contrast was obtained. Even here, the results were not comparable to any of those with copper radiation. Radiographs with cobalt radiation also failed to reveal the iron-rich particles at the high

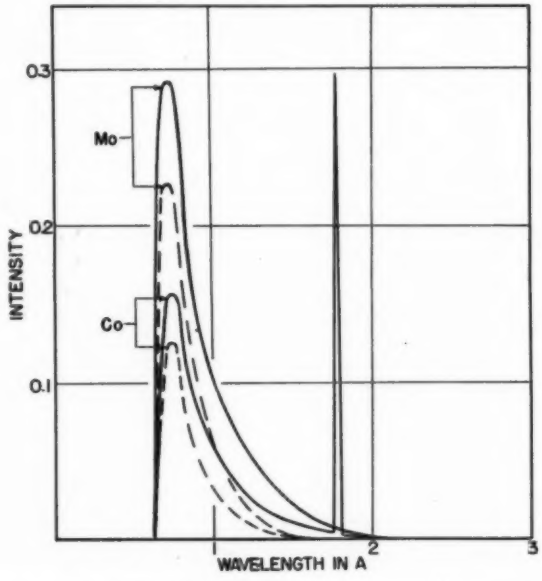


FIG. 8. Transmitted intensity curves for the two phases of the Al-Mn alloy for Mo and Co targets at 20 kv. (Particle thickness assumed = 0.01 cm. Scale expanded so that heights of the background peaks are approximately the same as in Fig. 7.)

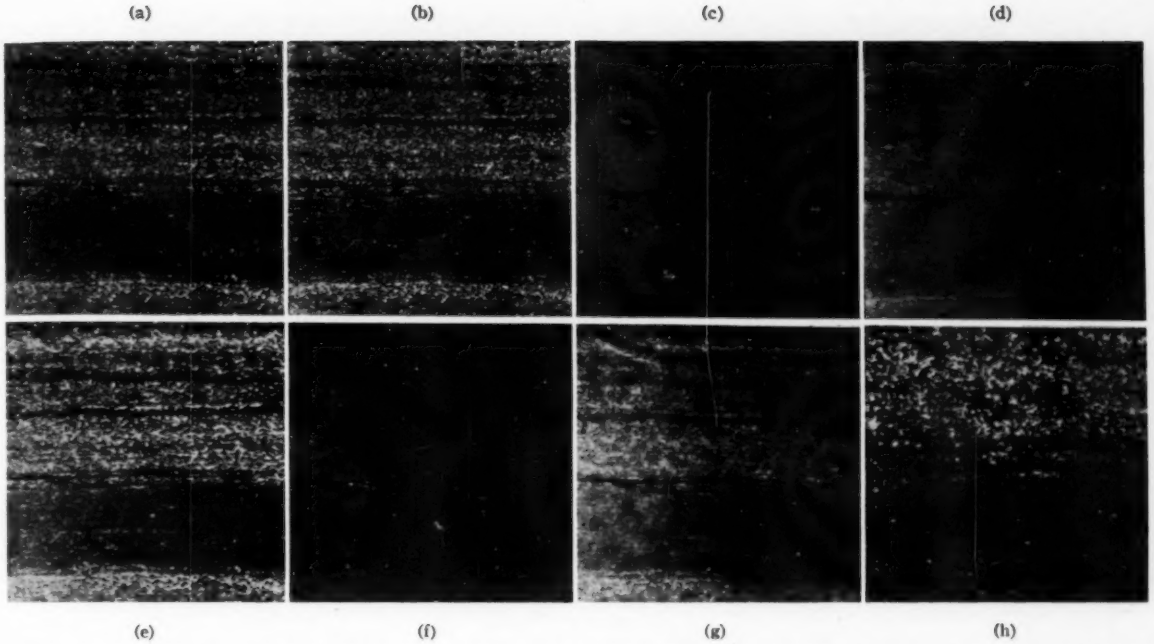


FIG. 9. Microradiographs of Cu-Zn-Fe-Si-Sn alloy. (a) to (g) at 50×; (h) at 100×. Cu-rich phase appears as dark background; Fe-rich phase appears as light grey to white equiaxed particles; Si-rich phase appears as black stringers.
(a) Cu radiation, 42 kv (b) Cu radiation, 20 kv (c) Mo radiation, 42 kv
(d) Mo radiation, 30 kv (e) Mo radiation, 17 kv (f) Co radiation, 42 kv
(g) Co radiation, 17 kv (h) Cu radiation, 20 kv

voltages since again the negative absorption differential for the characteristic radiation offsets any effect produced by the continuous background in the narrow, highly sensitive region. As the voltage is reduced, the Co K intensity does not approach zero as rapidly as that for molybdenum and, therefore, it was necessary to go to even lower voltages with cobalt before any indication was found of the iron-rich particles. Thus, even at 17 kv with the cobalt tube the radiograph showed the iron particles only faintly and this radiograph is not comparable with any taken with copper, nor with molybdenum at the lower voltages.

The emission of characteristic rays varies according to the equation[7]

$$I=k\{(V/V_j)-1\}^{1.65},$$

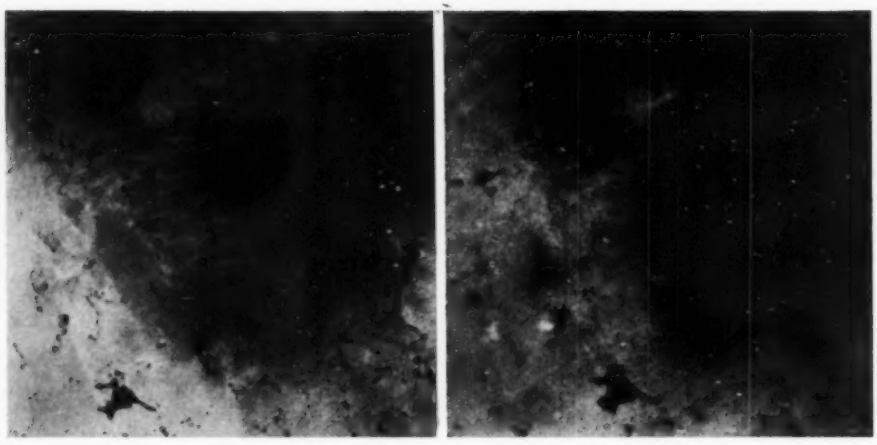
whereas the background radiation obeys the law $I=KZV$. This results in an increasing proportion of characteristic radiation at higher voltages. The absorption coefficient, except at absorption discontinuities, obeys the relation $\log \mu_m = A + n \log \lambda$ which effects an intensification of shorter wavelengths in the transmitted beam. There are thus two opposing effects, one favoring the predominance of the characteristic rays, the other favoring the short wave background. It is to be expected that with variation in voltage a maximum will be found for the effect of characteristic radiation.

This was not observed in the copper-tin or in the aluminum-manganese alloys, although a flattening out seemed to occur at lower voltages. With the iron-containing brass, however, a maximum did seem to exist and the contrast with copper radiation at 20–30 kv was slightly better than at either 42 or 17 kv.

II. Specimen-Target Distance

A considerable number of specimens were radiographed at specimen-target distances of 2 to 6 in. For fairly large particles, and particularly when the contrast was high, it was difficult to observe any deleterious effects even at the

[7] A. H. Compton and S. K. Allison, *X-Rays in Theory and Experiment* (D. Van Nostrand Co., Inc., 1935), p. 81.



(a) (b)

FIG. 10. Microradiographs of cartridge brass with Co radiation at 20 kv at 75×. Films of impurity are light grey against darker background of the brass matrix. Small voids are still darker than background.
(a) Target-specimen distance of 6 in. (b) Target-specimen distance of 2 in.

shortest distance. This was evident for the radiographs of the alloy of copper-zinc-tin (Table II). However, when the photographic contrast was low and when the particles were very small in one or more dimensions, the distance between target and specimen had a considerable effect upon the details observed. This is indicated in Fig. 10 which shows radiographs of a sample of cartridge brass in the cast condition. At a distance of 6 in. very fine interdendritic films of an impurity showed as a light tracery on a darker background. At a distance of 4 in. these films were somewhat fuzzy and had less contrast than they had at 6 in., while at 2 in. the detail of the films was completely eliminated; the small voids (black objects in the radiograph) were still readily discerned at 2 in. but were less sharp than at 4 or 6 in.

III. Inclusion Thickness and Magnification

These two items are closely linked with one another since the smallest size of inclusion which can be studied is dependent to some extent on the maximum useful magnification characteristic of the radiographic emulsion. Figure 9(h) is an enlargement to 200× of the same region as in Fig. 9(b). It is readily seen that a magnification up to 200× showed useful increase in detail. At 300× a general blurring occurred which rendered ineffectual the higher magnification. Aside from the useful magnification which can be obtained, the size of the particle which can be readily detected depends upon the absorption differential $(\mu_2-\mu_1)$. Among the iron particles visible in Fig. 9(h) can be seen a number which are of the order of 0.1 mm at 200× (about 0.00005 cm actual size), and as shown in Fig. 4 the absorption differential in this case is 2000 for the Cu K radiation. This corresponds to $I_2/I_1=0.9$ for the characteristic rays. Thus, with favorable operating conditions adequate contrast is obtained if $t_2 \geqslant \log 0.9/(\mu_2-\mu_1)$.

IV. Specimen Thickness

As was previously mentioned, the photographic contrast in the radiograph is given by $\exp[-(\mu_2-\mu_1)t_2]$. This should be entirely independent of specimen thickness. However, it is well known that in macroradiography, due to scattering effects, the observable detail does depend upon specimen thickness. In the extremely thin specimens used in microradiography it is hard to see how any such difficulty could occur. Nevertheless, experience gained from examination of a large number of specimens indicates that radiographs are slightly clearer when the specimen thickness is reduced. In cases where the inclusions penetrate through the full thickness of the specimen, the value t_2 of course depends upon the thickness of the specimen itself and contrast

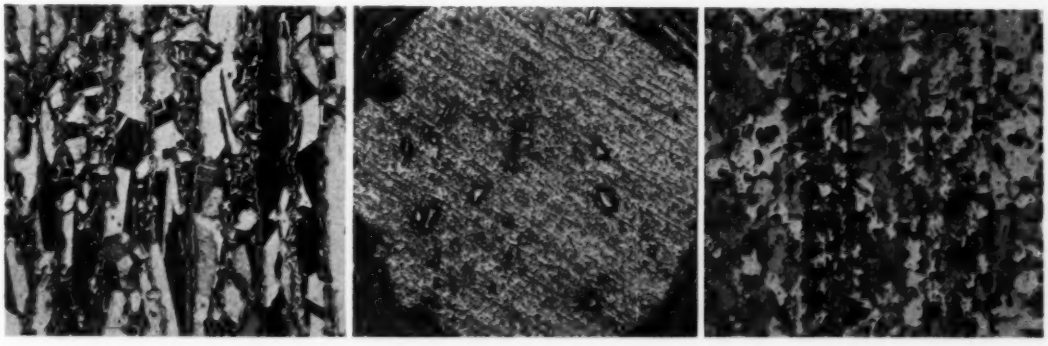
will increase with the specimen thickness. In the case of inclusions much smaller than the specimen thickness, there exists a possibility of a statistical averaging of the particles contained in different layers of the specimens, so that as a result the detail would be greatly obscured. In our experience, this type of difficulty has seldom occurred when the particles were extremely small, but has occurred for fairly large sizes. The specimen of 80 percent copper-10 percent tin-10 percent lead showed very poor detail when the gauge was double that finally used for Fig. 2. As a general rule, best results were obtained for specimen gauges less than 0.005 in. (See Fig. 11.)

CONCLUSIONS

From the results obtained here it can be stated that, in general, the use of characteristic rays with preselected wave-lengths is desirable for the detection of inhomogeneities such as inclusions or regions of secondary phases. For maximum sensitivity, however, the background radiation must not be ignored and it becomes necessary to adjust the tube voltage to suit the individual target element. Considerable latitude is permissible in the value of the tube voltage which makes it desirable to strike a careful balance between the economics of exposure time and the sensitivity required to bring out the inhomogeneities suspected to be present from micrographic or other studies. Usually a preliminary radiograph at the highest permissible voltage for a given target will serve to determine whether lower voltages are necessary. An estimate can be made of the optimum voltage range for a given target element by a study of the intensity distribution curves at various voltages. However, final judgment on this voltage range can only be dictated by previous experience with a variety of alloys.

The target-specimen distance will depend upon the type of x-ray tube and upon the inhomogeneities themselves. The shortest possible distance is desirable in order to minimize exposure times. Usually some prior knowledge of the size and composition of the inhomogeneities is already available from micrographic studies. This can be combined with previous experience to estimate a suitable target-specimen distance. In our experience most subjects yielded satisfactory radiographs at a distance of 4 in.; however, large inhomogeneities with high contrast can be studied successfully at distances as low as 2 in. In some cases of small inclusions or where the contrast is low, a distance of 6 in. may be necessary.

Perhaps the greatest future field of usefulness of the microradiographic method may lie in the identification of the elements contained in small inclusions observed in some alloys, or in the distribution of elements among the phases in ternary, quaternary, or even more complex systems. The method of procedure is readily deduced from the results with the iron-containing



(a) (b) (c)

FIG. 11. Photomicrographs for comparison with microradiographs.

(a) Cu-Zn-Sn at 300×. Elongated grains of alpha-phase with intergranular eutectoid of alpha- and gamma-phases.

(b) Al-Mn at 120×. Small inclusions of $Al\ Mn_6$ in Al matrix. Note paucity of inclusions compared to microradiograph.

(c) Cu-Zn-Fe-Sn-Si at 45×. Large grains of alpha- and beta-phases with small equi-axed particles of Fe-rich phase. Note paucity of inclusions compared to microradiograph.

brass. It can be readily seen here that, provided proper tube voltages are used, characteristic radiations in the narrow band between 1.377 and 1.739A will reveal the iron particles, whereas characteristic radiations which fall outside the band on either the short or long wave sides will not indicate their presence. Thus, the elements contained in the particles can be identified within limits by a study of the characteristic radiations capable of revealing them. However, for such a purpose the tube voltage would need to be carefully chosen to suit the target element; in many cases the whole effect can be vitiated by a poor choice of this important factor. The characteristic radiation should be strongly excited and yet the background peak must fall in a favorable region of the absorption differential curve. Subject to this consideration, the optimum operating voltage seems to be in the range 2.5 to 3.5 times the excitation potential. The economics of exposure time usually favor the higher part of this range.

This method has already found some success in this laboratory in identifying elements contained in inclusions and segregated particles. Further work is being done in this direction with regard to the target-voltage combinations necessary for the purpose.

ACKNOWLEDGMENT

The author is indebted to the Chase Brass and Copper Company for permission to publish this paper, and to Dr. D. K. Crampton, Director of Research, for his continued interest in the work. Thanks are also due to Messrs. B. R. Zimmerman and Harold Hughson for assistance in carrying out the radiographic and photographic work connected with the investigation.

Antenna Theory and Experiment

S. A. SCHELKUNOFF
Bell Telephone Laboratories, Inc., New York, New York

(Received September 20, 1943)

This paper presents: (1) a comparison between several approximate theoretical formulas for the input impedance of cylindrical antennas in the light of available experimental evidence; and (2) a discussion of the local capacitance in the vicinity of the input terminals, mathematical difficulties created by its presence, and methods of overcoming these difficulties. No exact solution of the antenna problem is available at present and so far it is impossible to set definite limits for errors which may be involved in various approximations. For this reason in appraising these approximations one is forced to rely on one's judgment and on experimental evidence. It is hoped that this paper will aid in correlating theory and experiment to the advantage of both.

DURING the past ten years several approximate formulas for the input impedance of antennas have been obtained by: (1) E. Siegel and J. Labus, for cylindrical antennas;[1] (2) E. Hallén, for cylindrical antennas;[2] (3) L. J. Chu and J. A. Stratton, for spheroidal antennas;[3] (4) S. A. Schelkunoff, for antennas of any shape.[4,5] The work of Hallén has been extended by Ronold King and F. G. Blake who published sets of impedance curves and tables[6] and obtained an explicit formula for the maximum

[1] E. Siegel and J. Labus, Hochfrequenz. und Elektroakustik **43**, 166–172 (1934).
[2] E. Hallén, Nova Acta (Uppsala) [4] **11**, 1–44 (1938).
[3] L. J. Chu and J. A. Stratton, J. App. Phys. 241–248 (1941).
[4] S. A. Schelkunoff, Proc. I.R.E. 493–521 (1941).
[5] S. A. Schelkunoff, *Electromagnetic Waves* (D. Van Nostrand Company, Inc., New York, 1943).
[6] Ronold King and F. G. Blake, Jr., Proc. I.R.E. 335–349 (1942). C. W. Harrison, Jr., J. App. Phys. **14**, 306–309 (1943).

resistance in the vicinity of $l=\lambda/2$, where l is the length of a vertical antenna just above ground, Fig. 1.

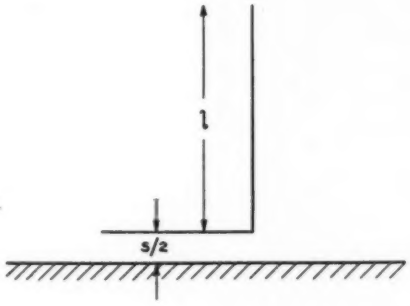


FIG. 1. An antenna above ground.

There is substantial disagreement between some of these formulas and while it is possible and legitimate to express opinions on the relative goodness of various methods of approximation, experiment will have to play an important role in the decision for some time to come. The point is that we have neither an exact solution nor an approximate solution with a remainder term such that we could set a quantitative upper bound to the error. Siegel and Labus rely to a large extent on a "physical" approximation; on the other hand, Hallén's approximations are largely, although not exclusively, mathematical. And yet from the quantitative point of view, Hallén's approximations are just as vague and uncertain as the physical approximations of Siegel and Labus.

By proving several exact theorems for biconical antennas, I have arrived at a method, based on approximate calculations of characteristic wave functions, which yields, I believe, good results for antennas whose length is at least moderately large compared with the diameter. The characteristic wave functions correspond to the normal modes of propagation and thus the theory conforms closely to our physical ideas of wave propagation. Even for large surface radiators this method of thought is useful. For example, I find that an approximate conductance of large spheres energized along the diameter is equal to $\pi a/60\lambda$, where a is the radius of the sphere and λ, the wave-length.[7] When plotted as a function of $2\pi a/\lambda$, this approximation is represented by a straight line about which the exact curve, obtained by Stratton and Chu, is wobbling up and down. Table I is a comparison for the integral values of $2\pi a/\lambda$.

TABLE I. The conductance of a spherical radiator.

$2\pi a/\lambda$	a/λ	The accurate value	The approximation
1	0.159	0.0125	0.0083
2	0.318	0.0168	0.0167
3	0.577	0.029	0.025
4	0.636	0.334	0.333
5	0.795	0.045	0.042

The same approximate expression is obtained for the conductance of a long cylinder whose radius is large compared with the wave-length; but even when the radius is equal to the wave-length, the approximation can hardly be distinguished from the accurate value obtained by numerical integration of the exact formula.* Nevertheless, I cannot say definitely how big or how small the error may be in any given case;† I can only say that in my judgment the method which I presented in references 4 and 5 will yield more accurate results than can be obtained from Hallén's formula and that the experimental evidence, meager though it is, supports my conclusion. Undoubtedly, additional experimental information is being accumulated and further experimental checks can be made by those possessing it.

Figure 2 shows the first maximum resistance $R_{\max.1}$ of a vertical wire above ground, Fig. 1; this maximum occurs when the length of the wire l is somewhat less than half wave-length. The experimental point corresponding to the lowest value of $l/2a$, where a is the radius of the antenna, was obtained by J. F. Morrison and P. H. Smith[8] by measuring the input resistance of a square tower made with steel girders. The

[7] See reference 5, page 476.

* Equation (164) of reference 4.

† For thin antennas it is easy enough to state "the order of magnitude" of the error; but in practical applications this would not mean much. Even though the errors are of the same order of magnitude, one approximation may turn out to be good and the other worthless.

[8] J. F. Morrison and P. H. Smith, Proc. I.R.E. 673–696 (1937).

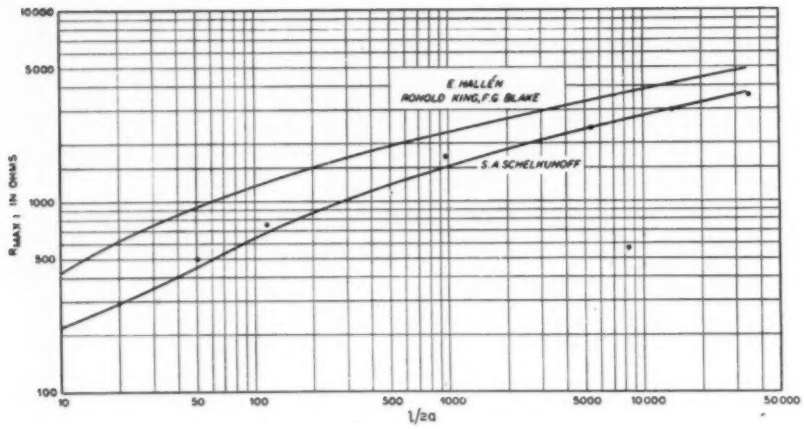


FIG. 2. Theoretical curves and experimental points for the first maximum resistance of a cylindrical antenna over ground as shown in Fig. 1; a is the radius of the antenna and λ, the wave-length.

remaining points were obtained by C. B. Feldman in 1934 and with his permission published[4] by me in 1941. Feldman's data were taken at a wave-length about equal to 18 meters; a substitution method was employed. The wave-length employed by Morrison and Smith was about 950 feet.

Theoretical curves have been calculated for the following conditions:

1. Ground is a perfect conductor.
2. Cylindrical radiators are perfect conductors.
3. The radiators are either infinitely thin hollow cylinders or their radii are so small that the "top capacitance" is negligible.
4. The distance between the input terminals is fairly large compared with the diameter of the cylinder so that the local capacitance which we may subdivide into the "base capacitance" and the "near-base excess capacitance" are negligible. While in Schelkunoff's theory these localized effects as well as the top capacitance can readily be taken into consideration, in Hallén's analysis they are entirely ignored.

In the case of thin wires, conditions (3) and (4) are usually met in practice. Except in the case of short antennas, their resistance is of secondary importance and it can readily be estimated by computing the loss on the assumption of the sinusoidal distribution of current in the antenna. No further refinement is necessary. The effect of ground may be more serious, especially when the current at the base of the antenna is large. Figure 2 exhibits a substantial divergence between two theoretical approximations.

For thin wires I obtained an approximate formula

$$R_{\max,1}=\frac{K_a^2}{400}-0.365K_a, \tag{1}$$

where K_a is the average characteristic impedance of the transmission line formed by the antenna and its image and

$$K_a=120\log\frac{2l}{a}-120$$

$$=276\ \mathrm{Log}_{10}\frac{2l}{a}-120\ \text{ohms}. \tag{2}$$

The corresponding expression obtained by Ronold King and F. G. Blake from Hallén's equations is

$$R_{\max,1}=\frac{(K_a+120)(K_a+68)}{400}\simeq\frac{K_a^2}{400}+0.475K_a. \tag{3}$$

The difference $0.84K_a$ increases as the antenna becomes thinner. In free space this difference is doubled.

While the difference between Hallén's and Schelkunoff's formula is particularly large in the vicinity of $l=\lambda/2$, it is substantial even for shorter antennas. In order to exhibit this difference let us compare each formula with the asymptotic formula (the formula for an infinitely thin antenna). As shown in Fig. 3 Hallén's curves practically coincide with the asymptotic curve until l/λ begins to approach 0.5; on the other hand, Schelkunoff's curves, Fig. 4, deviate almost everywhere in the interval $\lambda=4l$ and $\lambda=2l$ and show rather characteristic crossings. Hallén's curves are noticeably sharper. The curves in Figs. 3 and 4 represent the input resistance $\hat{R}_i$ of center-fed antennas in free space, that is, twice the input resistance of antennas energized as shown in Fig. 1.

The asymptotic expression for $\hat{R}_i$ is simply

$$\hat{R}_i \sim \frac{R_a(\beta l)}{\sin^2 \beta l}, \quad \beta=\frac{2\pi}{\lambda}=\frac{\omega}{v}, \tag{4}$$

where $R_a(\beta l)$ is the well-known "radiation resistance" referred to the current maximum and calculated on the assumption that the current distribution is sinusoidal.* This expression is

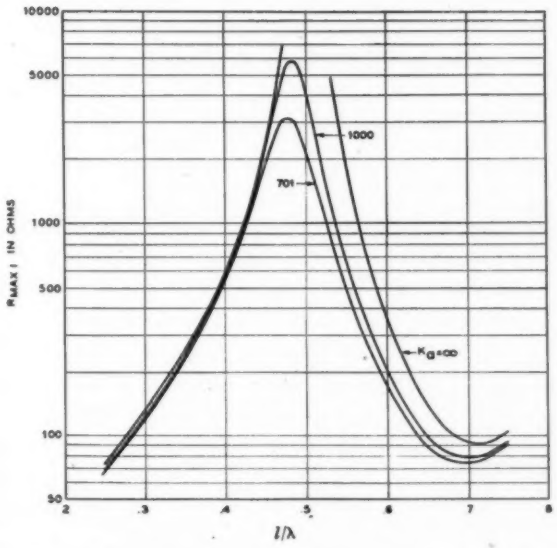


FIG. 3. Theoretical resistance curves (for an antenna in free space) obtained from Hallén's formula and the resistance curve for an infinitely thin ($K_a=\infty$) wire.

* Which is indeed the case if the antenna is infinitely thin.

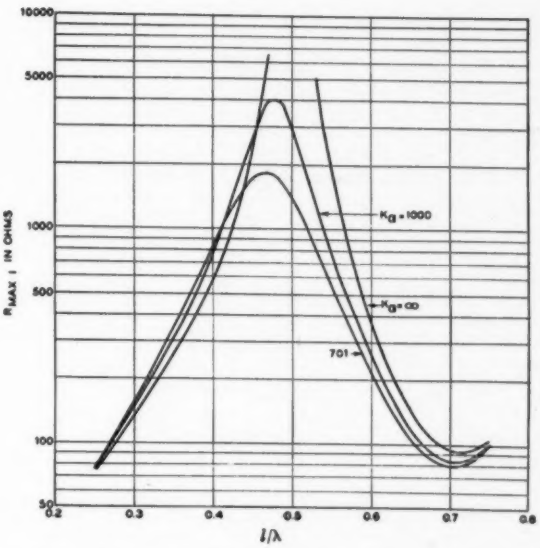


FIG. 4. Theoretical resistance curves (for an antenna in free space) obtained by Schelkunoff and the resistance curve for an infinitely thin ($K_a=\infty$) wire.

obtained at once from the energy theorem and the asymptotic properties of the antenna.

In their derivation of an approximate formula for the antenna impedance Siegel and Labus relied to a large extent on a physical approximation. One is apt to think that this kind of approximation is vaguer than the mathematical kind. In practice, however, this is not always the case since the remainder terms are often unknown. Siegel and Labus calculate the radiated power as usual from the sinusoidal distribution and then they introduce a series resistance, distributed along the antenna, so calculated that the total loss in the resistance equals the radiated power. Their formulas are

$$Z_i=R_i+iX_i=Z_A\left(1-\frac{i\alpha}{\beta}\right)\coth(\alpha+i\beta)l,$$

$$R_i=Z_A\frac{\sinh 2\alpha l-2\alpha l\dfrac{\sin 2\beta l}{2\beta l}}{\cosh 2\alpha l-\cos 2\beta l}, \quad \beta=\frac{2\pi}{\lambda},$$

$$X_i=-Z_A\frac{\sin 2\beta l+2\alpha l\dfrac{\sinh 2\alpha l}{2\beta l}}{\cosh 2\alpha l-\cos 2\beta l}, \tag{5}$$

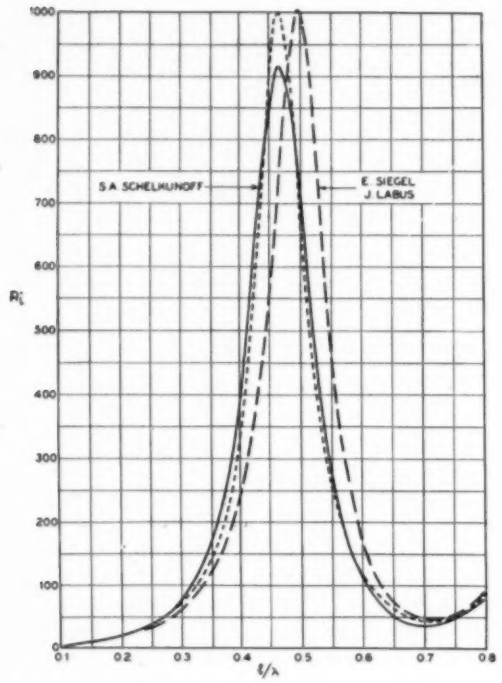


FIG. 5. The input resistance of a cylindrical antenna (in the arrangement shown in Fig. 1) *versus* the length in wave-lengths; $K_a=701$ ohms. The unlabeled curve is obtained when the Siegel and Labus formula is corrected for a reactive effect as explained in the text.

$$Z_A \simeq 120\left(\log\frac{l}{a}-1-\tfrac{1}{2}\log\frac{2l}{\lambda}\right),$$

$$\alpha l=\frac{R_a(\beta l)}{Z_A\left(1-\dfrac{\sin 2\beta l}{2\beta l}\right)}.$$

Figures 5 and 6 show these resistance and reactance characteristics as functions of the antenna length in wave-lengths for $K_a=701$ ohms. The maximum resistance agrees quite well with the experimental curve but it occurs practically at $l=\lambda/2$ which is at variance with experience. It has occurred to me that the Siegel and Labus formula could be improved as follows. We calculate the complex power flow from the antenna, assuming a sinusoidal current distribution, and subtract from it the complex power flow from the generator into an open-circuited uniform transmission line whose characteristic impedance is Z_A as calculated by Siegel and Labus. Then we introduce a uniformly distributed series impedance such that the complex power flow into it is equal to this difference (Siegel and Labus consider only the real part of this correction term); and finally we calculate the correction to the propagation constant due to this added series impedance. The resulting formulas are the same as Eqs. (5) provided we replace β by β' where

$$\beta' l=\beta l+\frac{X'(\beta l)}{Z_A\left(1-\dfrac{\sin 2\beta l}{2\beta l}\right)},$$

$$X'(\beta l)=60\,\mathrm{Si}\,2\beta l+30(2\,\mathrm{Si}\,2\beta l-\mathrm{Si}\,4\beta l)\cos 2\beta l$$

$$+30\Big(2\log 2-1 \quad (6)$$

$$+\mathrm{Cin}\,2\beta l-\mathrm{Cin}\,4\beta l-\frac{\sin 2\beta l}{2\beta l}\Big),$$

$$\mathrm{Cin}\,x=\int_0^x\frac{1-\cos t}{t}dt=C+\log x-\mathrm{Ci}\,x.$$

The unlabeled curves in Figs. 5 and 6 are obtained from these modified Siegel and Labus formulas.

The input impedance of spheroids can be expressed in terms of spheroidal functions as done by Stratton and Chu. However, some of their numerical results are inaccurate for the following reasons. They have assumed that the distance between the input terminals is zero. Inevitably this leads to an infinite capacitance across the terminals (since the radius of the spheroid is finite there) and hence an infinite susceptance. They actually observed that their series for the susceptance was divergent; but they felt that all they had to do in the case of a finite gap was to take only the first few terms of the series. What actually happened was that in effect they have introduced an unknown susceptance across the terminals of their spheroidal antenna. Whether their curves actually correspond to some particular length of the gap is hard to tell; anything may happen. I am reasonably certain, however, that their curves for very thin spheroids are inaccurate when the distance between the input terminals is fairly large compared to the diameter. Thus for a

spheroid whose length is 0.8λ and is 7070 times as large as the diameter at the center, they give 425 ohms as the input resistance. For the same case I obtain 638 ohms and the asymptotic value which ignores the radius and the shape of the antenna is 574 ohms.

LOCAL CAPACITANCE BETWEEN THE INPUT TERMINALS

The principal function of the local capacitance between the input terminals of an antenna is—so it seems—to distress applied mathematicians. Experimental engineers are not worried by this capacitance because when it is negligible they are blissfully unaware of its existence and when it is not negligible they know it is present and find ways of appraising its effect.

For example, let us take a parallel pair one mile long and let it be terminated at the far end into some impedance. Let the distance between the wires be half an inch and the frequency at which we wish to obtain the impedance of the line be one megacycle. When connecting this

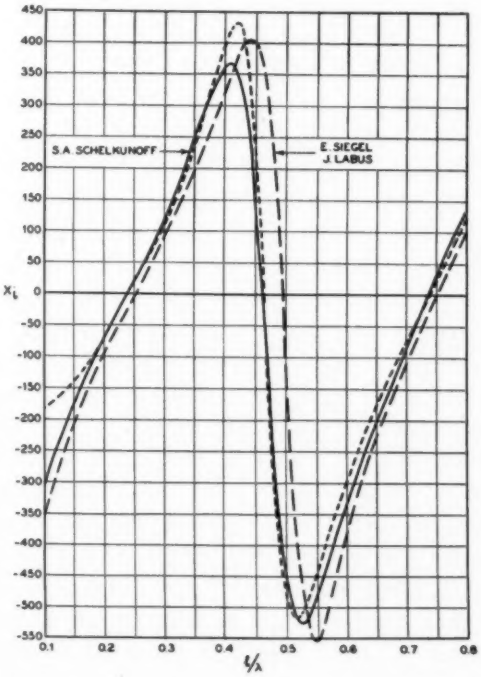


FIG. 6. The input reactance of a cylindrical antenna (in the arrangement shown in Fig. 1) *versus* the length in wave-lengths; $K_a = 701$ ohms. The unlabeled curve is obtained when the Siegel and Labus formula is corrected for a reactive effect as explained in the text.

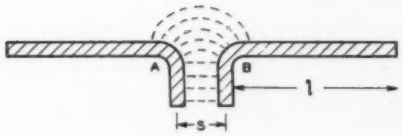


FIG. 7. An antenna connected to a parallel wire transmission line.

line to the measuring set, we shall probably spread the wires apart near the input end. One familiar with transmission lines would feel that the effect of this local change in the separation between the wires is negligible. If pressed, one could estimate this effect and obtain some figures to show how small it is. One would say that the separation between the wires right at the input end is not an important factor in the input impedance of a long line. Of course, it can be pointed out that this is not true since the ends of the wires could be brought so close together that the capacitance between them would become very large; so large, in fact, that it would be the principal factor in the impedance. However, this situation can occur only if the distance between the wires is small compared to their radii—a purely academic possibility in the case of "wires." On the other hand, this capacitive effect makes it generally impossible for the mathematician to use his favorite method of simplifying problems of this kind, that is, to make his generator a point source. In practice, the distance between the terminals is frequently unimportant; but in mathematics it is always important and unless properly treated may be a cause of serious errors.

Similarly, if an antenna is fed by a parallel wire transmission line, Fig. 7, the distance between the antenna "terminals" A, B will frequently be neither too large nor too small and its variation will not be an important factor in the antenna impedance. In the case of "fat" cylinders, however, the antenna "gap" may assume an important role, Fig. 8. If the cylinders are solid, there is the "base capacitance" approximately equal to $\epsilon\pi a^2/s$ (provided the radius of the cylinder is small compared with the wavelength). If the length s of the gap is small compared with the circumference, the capacitance per unit length between the cylinders in the vicinity of the gap (where the distance z

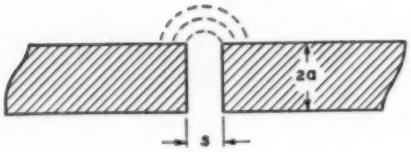


FIG. 8. The cross section of a cylindrical antenna of "large" radius.

from the central plane is small compared with a) is

$$C=\frac{\epsilon}{\pi z}\times 2\pi a=\frac{2\epsilon a}{z}; \tag{7}$$

but when z is large compared with a, the capacitance is more nearly

$$C=\frac{\pi\epsilon}{\log\,(2z/a)}. \tag{8}$$

The conditions in the vicinity of a narrow gap are rather different from those at some distance from it. At larger distances inductive effects are just as prominent as capacitive, and we have typical conditions for wave propagation; but in the vicinity of the gap, the capacitance predominates. It is possible to replace the antenna in Fig. 8 by the tapered antenna in Fig. 9(B) if in parallel with the latter we assume an "excess capacitance" equal to the difference between the actual near-base capacitance and the one in the modified antenna. The calculations can be carried out on the basis of *electrostatic* equations. The tapered antenna with conically shaped terminals has no singularity at the center and its impedance (or admittance) can be calculated from the equations given in references 4 and 5; then we add the admittance representing the excess capacitance.

In earlier work, including Hallén's, the gap was filled as shown in Fig. 9(A) to permit the charge to pass from one half of the antenna to the other half under the influence of the impressed electric intensity on the surface of the filled gap. Furthermore, Hallén assumed that the length of the gap is infinitely small. Thus, if he had carried out his calculations exactly, he would have obtained inevitably zero for the input impedance. In that case his series for the input current would have to diverge by the very nature of the physical problem he set out to solve. By

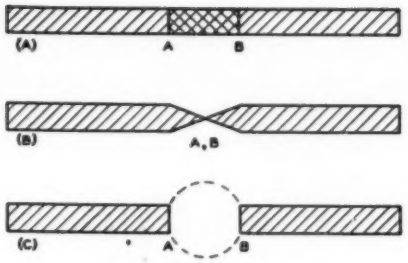


FIG. 9. Illustrating three methods of mathematical treatment of the antenna "gap."

assuming that the antenna current flows along the axis rather than on the surface of the antenna, Hallén removed one cause of the otherwise inevitable divergence of the series for the admittance although he does not appear to have realized it. In a sense this tacit assumption is a redeeming feature of his analysis since in effect the singular gap of Fig. 9(A) is replaced by the non-singular gap of Fig. 9(B). In any mathematical analysis based strictly on the arrangement 9(A) the length of the gap must be kept finite; and unless the length of the gap is fairly large compared with the diameter, the antenna current in the vicinity of the gap must be kept on the surface even for very thin wires. The results should then be interpreted in relation to the actual method of energizing the antenna.

In the antenna analysis used in references 4 and 5 the gap is treated in accordance with Figs. 9(B) and 9(C). This method conforms to actual methods of feeding antennas and provides an ample opportunity for including the effects of insulators and other environmental features. In this method the mathematics is adapted to the physics of the problem; and, in the light of this method, it is easy to interpret the results obtained for the gap in Fig. 9(A).

Although I have stressed possible inaccuracies of Hallén's formula, I do not wish to leave an impression that his basic idea of using an integral equation for solving the antenna problem is unsound. On the contrary, if used in conjunction with the method based on characteristic wave functions, it may yield valuable information. But first of all Hallén's method of successive approximations has to be revised; one such revision is given in the following paper by Miss Marion C. Gray.

A Modification of Hallén's Solution of the Antenna Problem

MARION C. GRAY
Bell Telephone Laboratories, Inc., New York, New York
(Received October 12, 1943)

An alternative formula for the input impedance of a cylindrical antenna is derived from Hallén's integral equation. It is shown that the introduction of a variable parameter $Z(z)$ in place of Hallén's $\Omega=\log(4l^2/a^2)$ modifies the numerical results considerably, and leads to much better agreement with experimental evidence.

THIS paper may be regarded as a supplement to a companion paper by S. A. Schelkunoff.[1] The latter has discussed the various methods that have been used in work on the antenna problem and has compared the theoretical results with the available experimental evidence. Since all the methods involve approximations, both physical and mathematical, it seems that the best solution must be chosen simply by the pragmatic test of agreement with experiment. From this standpoint the integral equation method of Hallén[2] as applied to a thin cylindrical antenna appears to give much less accurate results than the more physical methods of Schelkunoff or of Siegel and Labus. Yet the integral equation in itself is not invalid and it seems of interest to attempt to modify Hallén's solution, to bring it more nearly into line with the work of the other investigators.

Hallén's integral equation for the current in an antenna may be written[3] (in m.k.s. units)

$$\int_{-l}^{l}\frac{I(\xi)e^{-i\beta r}}{r}d\xi=\frac{-i4\pi}{\eta}(A\cos\beta z+V_0\sin\beta|z|), \tag{1}$$

where $2V_0$ is the voltage on the tuning circuit of a center-fed antenna of length $2l$, A is an arbitrary constant which is determined to satisfy the boundary condition $I(l)=0$, η is the intrinsic impedance of free space, $\eta=120\pi$, approximately, and

$$r^2=(z-\xi)^2+a^2, \tag{2}$$

where a is the antenna radius. Hallén's solution consists of expanding the current as an asymptotic series in powers of a parameter $1/\Omega$ where

$$\Omega=\log(4l^2/a^2). \tag{3}$$

Now while in theory the expansion might be continued to as many terms as desired, in practice the integrations involved become very onerous, and in fact Hallén has used only the first two terms of the expansion. Since no remainder term is available, the accuracy of the expansion cannot be checked, and it is obvious that a solution

$$I(z)=(f_1(z)/\Omega)+(f_2(z)/\Omega^2)$$

might equally well have been obtained in the form

$$I(z)=[F_1(z)/(\Omega+\alpha)]+[F_2(z)/(\Omega+\alpha)^2].$$

Further, although the two forms will tend to coincide when Ω is sufficiently large, within the practical range of values of Ω the second form may be much better than the first.

Hallén's choice of the parameter Ω is arbitrary, and somewhat unfortunate both from a physical and from a mathematical viewpoint. He writes

$$\int_{-l}^{l}\frac{d\xi}{r}=\log\frac{4l^2}{a^2}+\log\frac{l^2-z^2}{l^2}$$
$$+\log\frac{\{l-z+[(l-z)^2+a^2]^{\frac{1}{2}}\}\{l+z+[(l+z)^2+a^2]^{\frac{1}{2}}\}}{4(l^2-z^2)}$$

and then chooses for Ω the value of the integral at $z=0$. But this is the maximum value of the

[1] S. A. Schelkunoff, J. App. Phys. **15**, 54 (1944). References to the literature will be found in this paper.
[2] E. Hallén, Nova Acta (Uppsala) [4] **11**, 1–44 (November, 1938).
[3] It may be verified very simply by using the equations on pages 140 and 142 of *Electromagnetic Waves* by S. A. Schelkunoff (D. Van Nostrand Company, Inc., New York, 1943).

integral, and although Hallén himself does not attach any physical significance to his parameter, it may be regarded as proportional to the impedance of a transmission line. Thus his choice of Ω is equivalent to using the maximum value of the impedance in a non-uniform line to determine the current, whereas the average value would be a better choice.

From the mathematical point of view it also seems more logical to write

$$\int \frac{I(\xi)e^{-i\beta r}}{r}d\xi = \int \frac{[I(\xi)-I(z)]\cos\beta r}{r}d\xi + I(z)\int \frac{\cos\beta r}{r}d\xi - i\int \frac{I(\xi)\sin\beta r}{r}d\xi \quad (4)$$

rather than Hallén's equation

$$\int \frac{I(\xi)e^{-i\beta r}}{r}d\xi = \int \frac{I(\xi)e^{-i\beta r}-I(z)}{r}d\xi + I(z)\int \frac{d\xi}{r};$$

that is, it is preferable to pick out as parameter that part of the integral which does not converge as $a\to 0$. The integrals involving $I(\xi)$ on the right-hand side of (4) now converge at $a=0$ and in them r may be replaced by $|z-\xi|$. Further, a new parameter has been introduced in place of Hallén's Ω, namely,

$$Z(z)=\frac{\eta}{2\pi}\int_{-l}^{l}\frac{\cos\beta r}{r}d\xi, \quad (5)$$

where the constant multiplier $\eta/2\pi=60$ has been included so that $Z(z)$ will be dimensionally equivalent to a wave impedance. It must be admitted that the choice of a real function $Z(z)$ in (5) rather than the complex one which would be obtained if $\cos\beta r$ were replaced by $e^{-i\beta r}$ in the integral is again arbitrary, and is influenced by the fact that the other writers have used real wave impedances.

Introducing $Z(z)$ into the original integral Eq. (1) leads to the equation

$$I(z)Z(z)=-\frac{\eta}{2\pi}\int_{-l}^{l}\frac{I(\xi)-I(z)}{|z-\xi|}\cos\beta(z-\xi)d\xi + \frac{i\eta}{2\pi}\int_{-l}^{l}\frac{I(\xi)\sin\beta|z-\xi|}{|z-\xi|}d\xi - 2i[A\cos\beta z+V_0\sin\beta|z|]. \quad (6)$$

In this form the equation may be solved by a method of successive approximations similar to that used by Hallén. Thus, writing

$$I_0(z)=\frac{-2i}{Z(z)}(A\cos\beta z+V_0\sin\beta|z|), \quad (7)$$

$$I_1(z)=\frac{-\eta}{2\pi Z(z)}\int_{-l}^{l}\frac{I_0(\xi)-I_0(z)}{|z-\xi|}\cos\beta(z-\xi)d\xi + \frac{i\eta}{2\pi Z(z)}\int_{-l}^{l}\frac{I_0(\xi)\sin\beta|z-\xi|}{|z-\xi|}d\xi, \quad (8)$$

with similar formulas for $I_2(z)$, etc., it can be shown that

$$I(z)=I_0(z)+I_1(z)+I_2(z)+\cdots$$

is a solution of (6) in any region where the series converges. Substituting for I_0 in (8) its value from (7) the value of I_1 is found to depend on various integrals of which

$$\int \frac{\cos\beta(z-\xi)}{|z-\xi|}\left(\frac{\cos\beta\xi}{Z(\xi)}-\frac{\cos\beta z}{Z(z)}\right)d\xi$$

may be taken as typical. Since this integral has no singularity at $z=\xi$ and since $Z(z)$ is a slowly varying function of z throughout most of the range $|z|\leqq l$, it is possible to assume that both $Z(z)$ and $Z(\xi)$ may be replaced by an average value

$$\bar{Z}=\frac{1}{2l}\int_{-l}^{l}Z(z)dz \quad (9)$$

and taken outside the integral sign. Proceeding in this way with the other integrals, and retaining only the first two terms in the current

series leads to the solution

$$I(z)=\frac{-2i}{Z(z)}[A\cos\beta z+V_0\sin\beta|z|]$$
$$-\frac{2i}{Z(z)\bar{Z}}[AF_1(z)+V_0G_1(z)], \quad (10)$$

where

$$F_1(z)\equiv f_1(z)+if_2(z)$$
$$=\frac{\eta}{2\pi}\int_{-l}^{l}\frac{\cos\beta(z-\xi)}{|z-\xi|}(\cos\beta z-\cos\beta\xi)d\xi$$
$$+\frac{i\eta}{2\pi}\int_{-l}^{l}\frac{\cos\beta\xi\sin\beta|z-\xi|}{|z-\xi|}d\xi,$$

$$G_1(z)\equiv g_1(z)+ig_2(z)$$
$$=\frac{\eta}{2\pi}\int_{-l}^{l}\frac{\cos\beta(z-\xi)}{|z-\xi|}(\sin\beta|z|-\sin\beta|\xi|)d\xi$$
$$+\frac{i\eta}{2\pi}\int_{-l}^{l}\frac{\sin\beta|\xi|\sin\beta|z-\xi|}{|z-\xi|}d\xi.$$

The constant A is now determined to make $I(l)=0$, namely,

$$A=-V_0\frac{\bar{Z}\sin\beta l+G_1(l)}{\bar{Z}\cos\beta l+F_1(l)},$$

and hence

$$I(z)=\frac{-2iV_0\bar{Z}}{Z(z)[\bar{Z}\cos\beta l+F_1(l)]}\Big\{-\sin\beta(l-z)$$
$$+\frac{1}{\bar{Z}}[-G_1(l)\cos\beta z-F_1(z)\sin\beta l$$
$$+G_1(z)\cos\beta l+F_1(l)\sin\beta z]$$
$$+\frac{1}{\bar{Z}^2}[G_1(z)F_1(l)-G_1(l)F_1(z)]\Big\}. \quad (11)$$

Actually the term in $\bar{Z}^{-2}$ should be added to the next term in the series [determined from $I_2(z)$], but it has been retained here because it is not negligible numerically compared with the preceding term and it seems advisable to make the current as nearly zero as possible at the ends of the antenna.

From (11) the modified input impedance formula is now obtained in the form

$$Z_i=\frac{2V_0}{I(0)}=-iZ(0)\frac{\bar{Z}\cos\beta l+F_1(l)}{\bar{Z}\sin\beta l+U(l)}, \quad (12)$$

where

$$U(l)=G_1(l)+F_1(0)\sin\beta l-G_1(0)\cos\beta l$$
$$+\frac{1}{\bar{Z}}[G_1(l)F_1(0)-G_1(0)F_1(l)]. \quad (13)$$

The various functions involved in this solution are determined by straightforward integration in terms of the cosine and sine integrals. The values may be summarized as follows:

$$Z(z)=60\Big\{\log\frac{[(l+z)^2+a^2]^{\frac{1}{2}}+(l+z)}{[(l-z)^2+a^2]^{\frac{1}{2}}-(l-z)}$$
$$-\mathrm{Cin}\,\beta(l+z)-\mathrm{Cin}\,\beta(l-z)\Big\},$$

$$\bar{Z}=120[\log(2l/a)+\log 2$$
$$-\mathrm{Cin}\,2L-(\sin 2L/2L)],$$

$$L=2\pi l/\lambda=\beta l,$$

$$Z(0)=120[\log(2l/a)-\mathrm{Cin}\,L],$$

$$f_1(z)=30\cos\beta z[\mathrm{Cin}\,2\beta(l+z)+\mathrm{Cin}\,2\beta(l-z)$$
$$-2\,\mathrm{Cin}\,\beta(l+z)-2\,\mathrm{Cin}\,\beta(l-z)]$$
$$+30\sin\beta z[\mathrm{Si}\,2\beta(l-z)-\mathrm{Si}\,2\beta(l+z)],$$

$$f_2(z)=30\cos\beta z[\mathrm{Si}\,2\beta(l+z)+\mathrm{Si}\,2\beta(l-z)]$$
$$+30\sin\beta z[\mathrm{Cin}\,2\beta(l+z)-\mathrm{Cin}\,2\beta(l-z)],$$

$$g_1(z)=30\cos\beta z[2\,\mathrm{Si}\,2\beta z-\mathrm{Si}\,2\beta(l+z)$$
$$-\mathrm{Si}\,2\beta(l-z)]+30\sin\beta z[\mathrm{Cin}\,2\beta(l-z)$$
$$-\mathrm{Cin}\,2\beta(l+z)+2\,\mathrm{Cin}\,2\beta z-2\,\mathrm{Cin}\,\beta(l+z)$$
$$-2\,\mathrm{Cin}\,\beta(l-z)+4\log[(l+z)/z],$$

$$g_2(z)=30\cos\beta z[\mathrm{Cin}\,2\beta(l+z)+\mathrm{Cin}\,2\beta(l-z)$$
$$-2\,\mathrm{Cin}\,2\beta z]+30\sin\beta z[\mathrm{Si}\,2\beta(l-z)$$
$$-\mathrm{Si}\,2\beta(l+z)+2\,\mathrm{Si}\,2\beta z],$$

where the values of $g_1(z)$ and $g_2(z)$ are given for

$z>0$ and

$$\text{Cin } z=\int_0^z \frac{1-\cos u}{u}du=C+\log z-\text{Ci } z.$$

Equation (12) has been used to compute the impedance curves of Figs. 1 and 2. To facilitate comparison with the curves of Schelkunoff's paper, the input resistance and reactance have been drawn as functions of l/λ for values of $\log(2l/a)$ corresponding to the values $K_a=701$, 1000 used in his Figs. 3 to 6. The curves obtained from (12) agree so closely with Schelkunoff's, especially for the larger K_a, that it is not feasible to plot them together. For the smaller value of l/a, the original Hallén formulas are plotted as dashed curves.

In computing from (12) the second-order terms in $U(l)$ have been retained throughout. That these make a difference in the input resistance values may be shown by comparing Schelkunoff's formula for $R_{i,\max}$ with that obtained from (12). From the complete formula the maximum resistance for a half-wave antenna above perfect ground is approximately

$$R_{i,\max}=(K_a^2/400)-0.33K_a,$$

whereas if the second-order terms in $U(L)$ are neglected, it is

$$R_{i,\max}=(K_a^2/400)-0.55K_a.$$

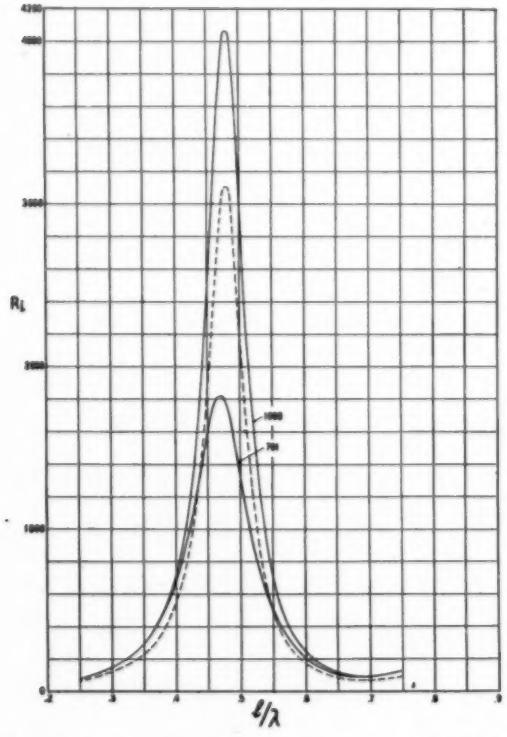


FIG. 1. The input resistance of a center-fed cylindrical antenna from the modified Hallén formula. The dashed curve is from the original Hallén formula for $K_a=701$.

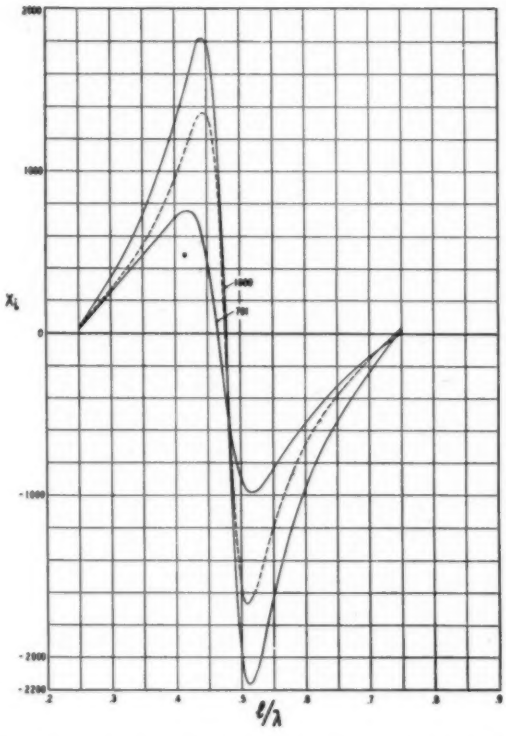


FIG. 2. The input reactance of a center-fed cylindrical antenna from the modified Hallén formula. The dashed curve is from the original Hallén formula for $K_a=701$.

It should be emphasized in conclusion that the solution of Hallén's integral equation presented here makes no claim to be *the* correct solution. The series for $I(z)$ is still an asymptotic series and it is conceivable that a different choice of parameter would prove even better. The introduction of the variable parameter $Z(z)$ and its average value $\bar{Z}$ has, however, modified Hallén's original solution considerably and has given numerical results in much better agreement with the available experimental evidence. The importance of Hallén's method lies in the fact that the next term of the asymptotic expansion can be found by straightforward

methods (though numerical methods of integration would have to be used). It is possible that computation of this next term would serve to decide between Hallén's expansion and the present one, if the numerical values were radically different.

Exploring Stress-Strain Relations of Isotropic Plastic Solids*

W. PRAGER
Brown University, Providence, Rhode Island
(Received November 5, 1943)

Combined torsion and tension of thin-walled tubes constitutes one of the few testing arrangements in which a fairly general state of uniform stress can be realized without too great experimental difficulties. A manner of representing graphically the results of such tests is used to discuss, in geometrical terms, some stress-strain relations of the mathematical theory of plasticity. It is shown that the theories of B. de Saint-Venant, M. Lévy, and R. v. Mises lead to physically unacceptable conclusions unless the material is supposed to be rigid as long as the stresses have not reached the yield limit. The predictions of the theories of L. Prandtl, E. Reuss, H. Hencky, and A. Nádai are compared with the results of experiments on mild steel; the possible effects of viscosity and strain hardening are discussed.

1. INTRODUCTION

SETTING out to explore the stress-strain relations of a given plastic material, one should always keep in mind that it is impossible to measure the local values of stress and strain at an arbitrarily assigned point of the test specimen. Only certain stress resultants and certain local deformations are accessible to measurement. For example, in a tension test the quantities actually measured are tensile force and elongation; in a torsion test, torque and angle of twist. The empirical relations between such stress resultants and the corresponding global deformations will permit the establishment of reliable stress-strain relations only if fairly uniform distributions of stress and strain are set up in the test specimen. Torsion of solid bars and bending being thus excluded, combined tension and torsion of thin-walled tubes recommends itself as one of the few tests in which a fairly general state of uniform stress can be realized without too great experimental difficulties. In this paper, a manner of representing graphically the results of such tests is used to discuss, in geometrical terms, some well-known stress-strain relations of the mathematical theory of plasticity.

Consider a thin-walled tube in combined torsion and tension. Let R denote the mean radius of the tube (Fig. 1), δ the thickness of its wall, λ and θ the elongation and the angle of twist of the gauge length L, S the tensile force, and T the torque. The tensile stress σ and the

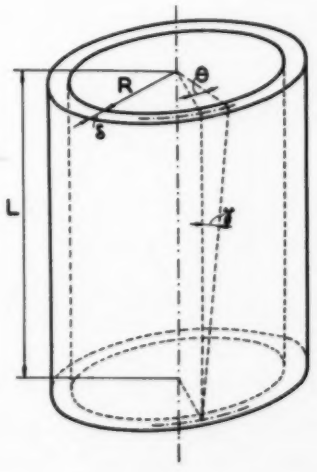


FIG. 1.

* This paper was presented to the Annual Meeting of the Society of Rheology, October 29, 1943.

shearing stress τ are then given by

$$\sigma = S/2\pi R\delta, \quad \tau = T/2\pi R^2\delta,$$

and the corresponding normal and shearing strains are

$$\epsilon = \lambda/L, \quad \gamma = \theta R/L.$$

If the material can be considered as incompressible, the logitudinal extension ϵ is accompanied by radial and circumferential contractions of the magnitude $\epsilon/2$. The maximum shearing strain thus is found to be

$$\gamma_{max} = [\gamma^2 + (3\epsilon/2)^2]^{\frac{1}{2}}, \tag{1}$$

while the maximum shearing stress equals

$$\tau_{max} = [\tau + (\sigma/2)^2]^{\frac{1}{2}}. \tag{2}$$

2. GRAPHICAL REPRESENTATION

Equations (1) and (2) suggest the following graphical representation of the states of stress and strain in a thin-walled tube under combined torsion and tension: In a rectangular coordinate system O, x, y (Fig. 2), the state of strain is

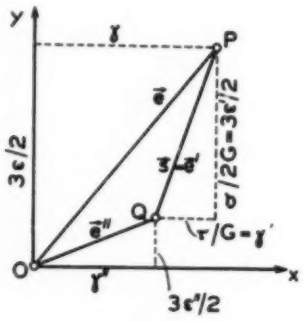


FIG. 2.

represented by the point P of the coordinates

$$x_P = \gamma, \quad y_P = 3\epsilon/2, \tag{3}$$

and the state of stress by the point Q of the coordinates

$$x_Q = x_P - \tau/G, \quad y_Q = y_P - \sigma/2G, \tag{4}$$

where G denotes the modulus of rigidity of the material under consideration.* According to (1) and (2), the vectors $\mathbf{e} = \mathbf{OP}$ and $\mathbf{s} = \mathbf{QP}$ have the magnitudes γ_{max} and τ_{max}/G, respectively. The vector $\mathbf{s}$ forms with the y axis the angle $\arctan 2\tau/\sigma$ which equals double the angle between the axial direction of the tube and the direction of the maximum normal stress. Similarly, the angle which the vector $\mathbf{e}$ forms with the y axis equals double the angle between the axial direction of the tube and the direction in which the unit extension assumes its greatest value. The vectors $\mathbf{s}$ and $\mathbf{e}$ thus describe completely the states of stress and strain in the tube under combined torsion and tension.

For the plastic materials considered here, the total strain (γ, ϵ) existing under a certain stress (τ, σ) can be decomposed into an elastic part (γ', ϵ'), disappearing upon unloading, and a permanent part (γ'', ϵ''), which subsists after unloading. The elastic strain components are connected with the stress components by means of

$$\gamma' = \tau/G, \quad \epsilon' = \sigma/E,$$

where E denotes Young's modulus. If, in accordance with the assumption made above, the material is considered as incompressible, $E = 3G$ and, consequently, $\epsilon' = \sigma/3G$. The components of the vector $\mathbf{QP}$ along the coordinate axes are thus seen to equal

$$\tau/G = \gamma', \quad \sigma/2G = 3\epsilon'/2. \tag{5}$$

Comparison of (5) to (3) shows that the vector $\mathbf{QP}$ represents the elastic strain in precisely the same way as the vector $\mathbf{e}$ represents the total strain. From this point of view the vector $\mathbf{QP}$ will be denoted by $\mathbf{e}'$ in the following; only if the vector $\mathbf{QP}$ is considered as representing the state of stress, will it be denoted by $\mathbf{s}$.

Since $\mathbf{e} = \mathbf{OP}$ represents the total strain and $\mathbf{e}' = \mathbf{QP}$ its elastic part, the vector $\mathbf{e}'' = \mathbf{OQ} = \mathbf{e} - \mathbf{e}'$ represents the permanent strain. In the case shown in Fig. 2 the vectors $\mathbf{e}$, $\mathbf{e}'$, and $\mathbf{e}''$ have different directions, i.e., the principal directions of the tensors of total strain, elastic strain, and permanent strain do not coincide.

3. PURE TORSION

Consider a thin-walled tube of a perfectly plastic material which yields under a constant maximum shearing stress $\tau_{max} = KG$, and behaves in an elastic manner whenever $\tau_{max} < KG$. Start-

* The materials considered here are assumed to behave in an elastic manner whenever the stresses fall short of those required to produce plastic deformations, i.e., during the initial stage of loading and during unloading. G refers to such elastic changes of state.

ing from a state of zero stress and strain, let this tube be deformed in a specified manner, represented by a given motion of the point P of the graphical representation discussed in the preceding section. Initially, both P and Q coincide at the origin O. As P starts on its path, Q will at first stay at O. Indeed, the permanent strain represented by $\mathbf{e}'' = \mathbf{OQ}$ vanishes as long as the stresses remain below the yield limit, i.e., as long as the magnitude of the vector $\mathbf{s} = \mathbf{QP}$ is smaller than K. In other terms, Q remains at O until P reaches the circle of the center O and the radius K. When P crosses this circle, plastic flow sets in. As this flow is assumed to occur under constant maximum shearing stress, $\tau_{max} = KG$, the vector $\mathbf{s}$ will retain the constant magnitude $\tau_{max}/G = K$ throughout the entire period of plastic flow. The point Q therefore follows P at the fixed distance K. This statement is, of course, not sufficient to permit the determination of the motion of Q, once the motion of P has been given. The precise manner in which Q follows P while preserving the distance $QP = K$, is defined by the stress-strain relations of the plastic material under consideration.

Before entering upon the discussion of the plastic behavior of a thin-walled tube under combined torsion and tension, let us discuss the simpler case of pure torsion. The lower half of Fig. 3 shows the (idealized) stress-strain diagram

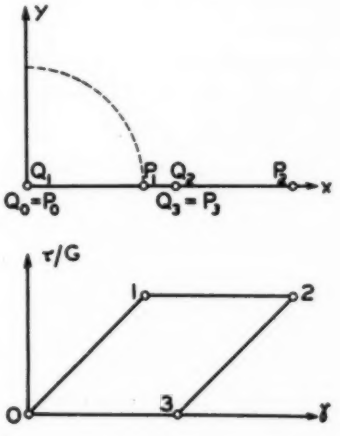


FIG. 3.

of a torsion test. The *dimensionless shearing stress* τ/G is plotted against the shearing strain γ; the portion *0–1* corresponds to the initial elastic behavior, the portion *1–2* to the plastic flow under constant shearing stress, and the portion *2–3* to the unloading. The abscissa of the point *3* represents the permanent shearing strain. In the terms of the graphical representation discussed above the same experiment would be described as follows: Initially the points P and Q coincide at the origin (P_0, Q_0 of the upper half of Fig. 3). As the tube is twisted, P moves along the positive x axis while Q stays at the origin as long as P remains inside the *yield circle*, i.e., the circle of the center Q_0 and the radius K. When P crosses this circle at P_1, Q starts to follow P at the fixed distance K, both P and Q moving along the x axis. If the tube is unloaded after P and Q have reached the positions P_2 and Q_2, respectively, the point Q remains at Q_2 since the permanent deformation represented by Q_0Q_2 remains unchanged during the process of unloading. As the stress is represented by the vector $\mathbf{s} = \mathbf{QP}$, the unloading is completed when the point P has reached the position $P_3 = Q_3 = Q_2$.

In order to visualize this motion of the points P and Q, one may think of these points as of two particles connected by an inextensible string of the length K. Initially, P and Q coincide at the origin and the string is slack. As P moves along the x axis,* Q remains at the origin until P reaches P_1, and the string becomes taut. Thereafter Q follows P at the fixed distance K. If the sense of motion of P is reversed, the string becomes slack again and Q stops.

The behavior of the test specimen in pure tension is obviously represented by the same mechanical model. A slightly more general manner of straining the tube lets the point P move along a straight line through O. For an isotropic material the principal axis of stress and strain can be expected to coincide during such a test. This means that the vectors $\mathbf{e}$ and $\mathbf{s}$ have the same direction and that the mechanical model again gives a correct representation of the behavior of the test specimen. In all cases discussed so far the vector $\mathbf{s} = \mathbf{e}'$ remains constant in direction and magnitude throughout the period of plastic flow, that is, the elastic strain

* In order to exclude inertia effects which would not have a counterpart in the plastic behavior of the twisted tube, this motion must be visualized as indefinitely slow.

components remain constant, and the change in total strain is entirely due to the change in permanent strain. In the following, such changes of strain will be referred to as *purely plastic changes of strain.*

4. COMBINED TORSION AND TENSION

As far as the behavior of a thin-walled tube under combined torsion and tension is concerned, a stress-strain relation amounts to a rule by which the motion of Q can be predicted when the motion of P is given. Three such rules will be considered in the following; they all appear as natural generalizations of the rules formulated above for some special cases.

a. Theories of Saint-Venant, Lévy, and Mises

A first rule governing the motion of Q can be stated as follows (see Fig. 4a). *As long as P stays*

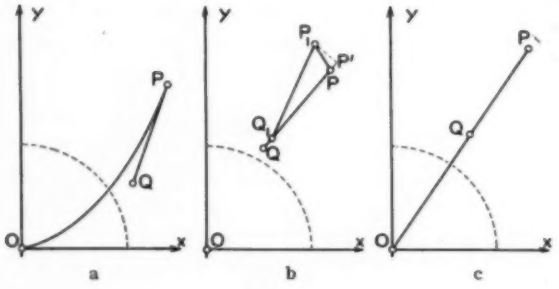


FIG. 4.

inside the yield circle, Q remains at O (elastic loading). When P is outside this circle, different rules apply according to the direction of motion of P. Through P draw the perpendicular p to QP. According to whether P moves towards the side of p on which Q is situated or towards the opposite side, elastic unloading or plastic flow will occur. In the case of plastic flow the vector $\mathbf{s} = \mathbf{QP}$ *remains tangent to the path of P at the instantaneous position of this point, and preserves the constant length K. In the case of elastic unloading, Q retains its position which, in the case of a subsequent loading, will play the same role as the origin O did for the first loading.*

The velocity vector of P represents the strain velocity in precisely the same way in which the vector **OP** represents the strain. The rule stipulating that **s** should have the direction of the tangent of the path of P, i.e., the direction of the velocity of P, amounts therefore to the statement that the stress components are proportional to the corresponding components of the strain velocity. The factor of proportionality is, however, not a constant characteristic of the material, but is at any instant determined by the condition that the stress components fulfill the yield condition, i.e., by the condition that the vector **s** has the magnitude K. The rule formulated above is thus seen to correspond to the stress-strain relations of the mathematical theory of plasticity inaugurated by B. de Saint-Venant[1] and M. Lévy.[2] The theory of R. v. Mises[3] differs from this only insofar as another yield condition is used instead of the condition of constant maximum shearing stress. K. Hohenemser and W. Prager[4] have given a slightly modified form of graphical representation which is suitable for materials following this yield condition.

b. Theory of Prandtl and Reuss

The mechanical model discussed in the preceding section suggests the following possibility of formulating a rule which determines the motion of Q when the motion of P is known. *In order to obtain the motion of Q when that of P is given, visualize P and Q as slowly moving particles connected by an inextensible string of the length K.*

At any instant the point Q will then either be at rest (when the string is slack) or move in the instantaneous direction of QP (when the string is taut). In the latter case a small increment of strain (PP_1 in Fig. 4b) can be thought of as consisting of two components: PP', along QP, and $P'P_1$, perpendicular to QP. The corresponding small displacement of the two particles and the connecting string can accordingly be thought of as consisting of a translation from the position QP to the position Q_1P', followed by a rotation

[1] B. de Saint-Venant, J. d. Math. Pures et App. (2) **16**, 308 (1871).
[2] M. Lévy, J. d. Math. Pures et App. (2) **16**, 369 (1871).
[3] R. v. Mises, Göttingen Nachrichten, math. phys. Kl., p. 582 (1913).
[4] K. Hohenemser and W. Prager, Zeits. f. angew. Math. u. Mech. **12**, 1 (1932).

about Q_1 which brings the system into the position Q_1P_1. The translation considered alone corresponds to a purely plastic change of strain in the sense defined at the end of Section 3. During the rotation the particle Q retains the position Q_1, that is, the permanent strain remains constant. The corresponding deformation may therefore be described as purely elastic.

The stress-strain relations represented by this mechanical model are those of the theory of L. Prandtl[5] and E. Reuss.[6]

c. Theory of Hencky and Nádai

There exists a third way of generalizing, in an equally natural manner, the rules which govern the motion of Q in the cases of pure torsion or pure tension. This third rule can be formulated as follows (see Fig. 4c). *As long as P stays inside the yield circle, Q remains at O (elastic loading). When P is outside this circle the cases of plastic flow and elastic unloading are to be distinguished in the way described in 4a. In the case of plastic flow, Q always lies on the straight line OP, between these two points and at the fixed distance K from P. In the case of elastic unloading, Q retains its position which, in the case of a subsequent loading, will play the same role as the origin O did for the first loading.*

A stress-strain relation corresponding to this rule has first been formulated by Hencky[7] who uses, however, another yield condition in place of the condition of constant maximum shearing stress. A. Nádai's[8] well-known three laws of yielding state essentially the same stress-strain relation.

5. EXPERIMENTUM CRUCIS

It seems worth while to compare the predictions of the three theories a, b, c to each other as well as to what little information is available concerning the mechanical behavior of isotropic plastic solids under combined stress. The predictions of the three theories agree with each other whenever the test specimen is strained so as to make P describe a straight line through the origin. In order to test the theories, more general ways of straining the test specimen must therefore be considered. The following type of testing has been suggested by E. Reuss:[6] the test specimen is first stressed in pure torsion until the yield limit is reached; thereafter the angle of twist is kept constant while the tube is stretched in the axial direction. In Fig. 5a,

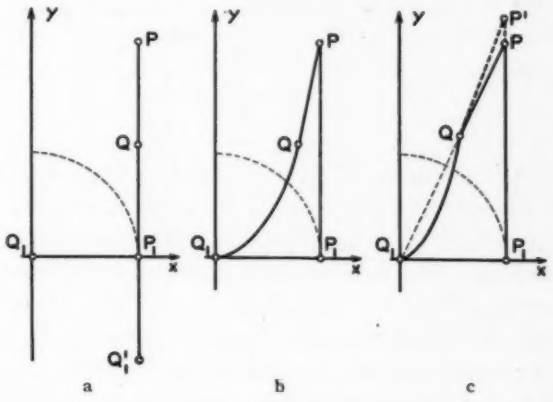


FIG. 5.

Q_1P_1P shows the path which the point P describes during this test.

It does not seem to be generally realized that in this case the theories of Saint-Venant, Lévy, and Mises lead to conclusions which it is difficult to accept. Indeed, according to these theories the vector **s**, which at first possesses the direction of the x axis, would suddenly swing into the position $Q_1'P_1$ as soon as P starts to move along the line P_1P. The corresponding sudden change of stress from pure torsion to pure tension may seem acceptable as the consequence of an equally sudden change in strain velocity from twisting to stretching. However, the shift of the vector **s** from the position Q_1P_1 to the position $Q_1'P_1$ reflects not only a change in stress but also a change in permanent strain which is represented by the displacement of Q from Q_1 to Q_1'. That such a change should occur while the total deformation, represented by P_1, remains constant, seems indeed unacceptable. There is but one way of avoiding this difficulty. Consider a series of plastic materials which behaves accord-

[5] L. Prandtl, "Spannungsverteilung in plastischen Körpern," Proc. 1st Int. Cong. Appl. Mech., p. 43 (1924).
[6] E. Reuss, Zeits. f. angew. Math. u. Mech. **10**, 266 (1930).
[7] H. Hencky, Zeits. f. angew. Math. u. Mech. **4**, 323 (1924).
[8] A. Nádai, *Plasticity, a Mechanics of the Plastic State of Matter* (McGraw-Hill Book Company, Inc., New York, 1931), p. 75.

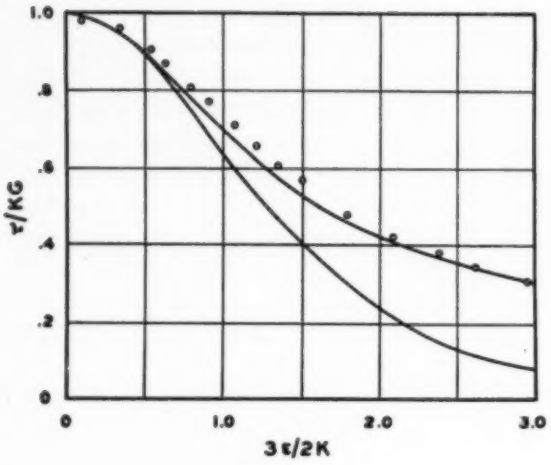


FIG. 6.

ing to the theory of Saint-Venant and Lévy, and which all have the same yield stress in shearing but different moduli of rigidity. The higher this modulus, the smaller is the constant K and, consequently, the nearer is Q_1' to Q_1. In the limit $K \to 0$, the sudden change of stress from pure tension to pure torsion would still exist but no longer the unacceptable sudden change of the permanent deformation. Now, $K=0$ characterizes a plastic material which cannot be deformed in an elastic manner, the modulus of rigidity being infinite. While such a material is frequently considered in presentations of the theories of Saint-Venant, Lévy, and Mises, this is generally justified by stating that the elastic deformations are small as compared with the plastic ones, and can therefore be neglected. Actually, these theories lead to unacceptable conclusions unless the plastic material is considered as rigid when under the influence of stresses below the yield limit.

It is easily seen that the theory of Prandtl and Reuss as well as that of Hencky and Nádai predict a continuous change of the stress as well as of the permanent strain. Figures 5b and 5c show the paths of Q according to these two theories; according to Prandtl and Reuss a tractrix is obtained, according to Hencky and Nádai a conchoide. In the first case the vector $\mathbf{s}$ approaches the vertical position more rapidly than in the second case. This means that the theory of Prandtl and Reuss predicts a more rapid decrease of torque with increasing extension than the theory of Hencky and Nádai. Experiments of this type permit therefore of deciding between these two theories. Figure 6 shows experimental results of K. Hohenemser[9] on tubes of mild steel; the quantity τ/KG is plotted against $3\epsilon/2K$, the lower curve corresponding to the theory of Prandtl and Reuss, the upper curve to that of Hencky and Nádai. The experimental points are seen to lie close to this latter curve. Subsequent experiments of K. Hohenemser and W. Prager[4] in which more general ways of straining were used, substantially confirmed this decision in favor of the theory of Hencky and Nádai.

6. PRESUMABLE EFFECTS OF STRAIN HARDENING AND VISCOSITY

It is to be expected that, to a certain extent, the effect of strain hardening may be represented by stipulating that during the plastic deformation the length of the vector $\mathbf{s}=\mathbf{QP}$ be increased in proportion to the length of the vector $\mathbf{e}''=\mathbf{OQ}$. This means that the vector $\mathbf{s}$ in Fig. 5b would approach the vertical position less rapidly than if its length were kept constant. On the other hand, an increase in length of the vector $\mathbf{s}$ in Fig. 5c would not influence the orientation of this vector which is always directed along OP. Consequently, strain hardening tends to reduce somewhat the difference between the predictions of the theory of Prandtl and Reuss and that of Hencky and Nádai.

A more general type of strain hardening lets the length of the vector $\mathbf{s}$ during the plastic flow depend not only on the length but also on the direction of $\mathbf{e}''$. In this way the anisotropy produced by strain can be represented. For theories of isotropic and anisotropic strain hardening the reader is referred to the literature.[10–12]

As regards the effects of viscosity, the stress-strain relations discussed so far may be considered as corresponding to indefinitely small rates of straining. For finite strain velocities the vector representing the stress appears as the

---

[9] K. Hohenemser, Zeits. f. angew. Math. u. Mech. **11**, 15 (1931).
[10] A. Nádai and E. A. Davis, J. App. Phys. **8**, 205 (1937).
[11] W. Prager, Mémorial d. Sci. Math., Fasc. 87 (Paris, 1927).
[12] W. Prager, *Theory of Plasticity*, mimeographed lecture notes (Brown University, Providence, Rhode Island, 1942).

sum of the vector **s**, considered so far, and a vector representing the viscous stress which may be taken proportional to the velocity of P. Consider, for instance, the vector **QP** of Fig. 5c corresponding to an indefinitely slow motion of P along P_1P'. If P moves with a finite velocity, a vector PP' proportional to this velocity must be added to **QP**, the total stress then being represented by the vector **QP'**.

7. CONCLUSION

Up to a fairly recent past, engineering research in plasticity was almost exclusively concerned with the yield condition. In fact, engineers were interested more in the limits of the elastic range than in the phenomena occurring beyond this range. The desire to obtain a basis for an adequate stress analysis in partly overstrained machine parts and the necessity of developing a better understanding of numerous technological processes have changed this situation. While many rival theories of plasticity have been developed by the theoreticians, experimental research has not kept pace with these theoretical developments. Considering the care with which the comparatively small differences between the more adequate yield conditions have been explored, it is surprising that the stress-strain relations valid in the plastic range have received so little attention. Research in this field has largely been restricted to cases of purely plastic changes of strain (see the definition at the end of Section 3).[13–15] In those cases the three mathematical theories of plasticity discussed in the present paper furnish identical predictions. More general ways of straining should therefore be used in exploring the stress-strain relations of plastic materials.

---

[13] W. Lode, Forschung. a. d. Gebiete d. Ing., No. 303 (Berlin, 1928).

[14] M. Roš and A. Eichinger, *Versuche zur Klärung der Bruchgefahr* (Diskussionsbericht No. 34 der Eidgen. Materialprüfungsanstalt, Zürich, 1929).

[15] G. I. Taylor and H. Quinney, Phil. Trans. Roy. Soc. **A230**, 323 (1931).

Studies in Three-Dimensional Photoelasticity
Stress Concentrations in Shafts with Transverse Circular Holes in Tension
Relation Between Two- and Three-Dimensional Factors*

MAX M. FROCHT†
Photoelastic Laboratory, Carnegie Institute of Technology, Pittsburgh, Pennsylvania
(Received October 8, 1943)

Part I. General problems of technique arising in three-dimensional photoelastic analysis by means of the frozen stress pattern method are discussed. These include suggestions for loading, slicing, improvement of boundary visibility, a discussion of time stresses, and methods for the determination of the fringe order and fringe value in a model with a static or frozen stress pattern. A number of instructive three-dimensional stress patterns are shown. Part II. Stress concentration factors for shafts with ratios of r/d equal to 0.0666, 0.323, and 0.164 are determined, r being the radius of the hole, D the diameter of the shaft, and $d=D-2r$. Part III. A relation is established between two- and three-dimensional factors of stress concentration for a range of r/d which does not exceed 0.5, which is believed to cover the most practical cases. It is shown that the three-dimensional factors are somewhat greater than the two-dimensional factors for the same ratio of r/d.

PART I

Introduction

IN this paper we report the results of a photoelastic investigation of the stress concentrations produced around open holes in circular shafts which are subjected only to tensile forces parallel to the axis of the shaft. This problem is not only of scientific interest but is also of practical importance. Open holes in shafts often form a part of the lubricating system of a machine. The oil holes in crankshafts are one example. A knowledge of the stress concentrations around such holes in shafts will also aid in establishing a relation between two- and three-dimensional factors of stress concentrations, since the corresponding two-dimensional problems have already been successfully solved. That this relationship is of practical value is evidenced by the fact that considerable research work in that direction has already been done by means of strain gauges and fatigue tests.[1] The development of three-dimensional photoelasticity makes

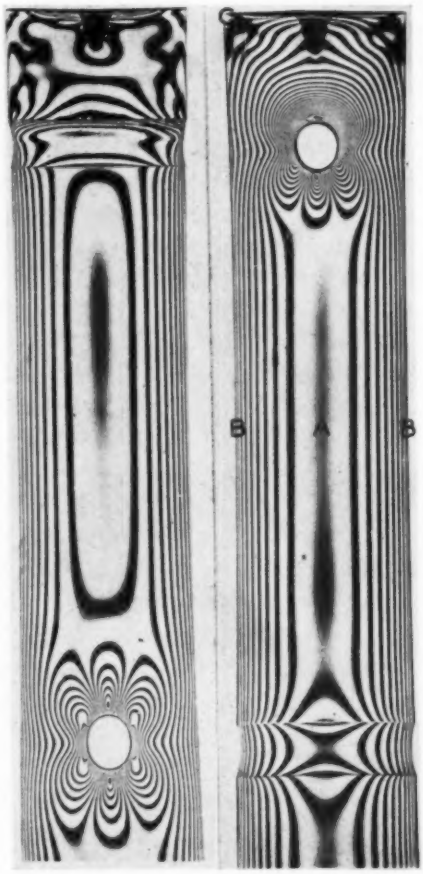


FIG. 1. Frozen stress pattern of a circular shaft with a transverse hole showing that the stresses in the shank are symmetrically distributed about the axis. $D=1.069''$; $r=0.132''$; $r/d=0.164$ (initial dimensions); $P=22.15$ lb.; f (nominal) $=1.42$ p.s.i. shear; $N=9.4$ fringes.

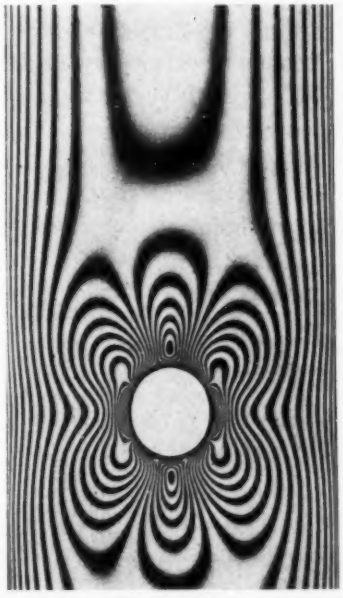
* This work is from a forthcoming book on photoelasticity to be published by John Wiley & Sons, Inc., and no part of it is to be reprinted without written permission from the author.

† Associate Professor of Mechanics in charge of Photoelastic Laboratory, Carnegie Institute of Technology, Pittsburgh, Pennsylvania.

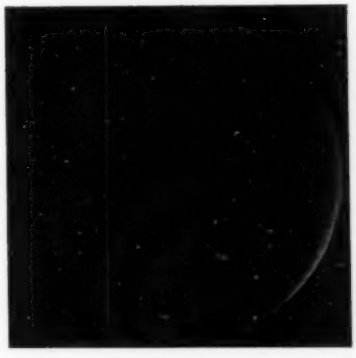
[1] See R. E. Peterson and A. M. Wahl, J. App. Mech. **3**, No. 1 (March, 1936).

it possible to study this problem by the optical method, with a reasonable expectation of greater accuracy. The holes were all circular and transverse to the axis of the shaft. A search of the literature fails to show any previous work on this particular problem.

In the course of this investigation certain problems of technique were encountered and solved either wholly or partially and it is believed that these technical aspects will be of value to other workers in the field. In Part I of this paper these general problems of technique are described. In Part II we deal with the specific problem of stress concentrations.

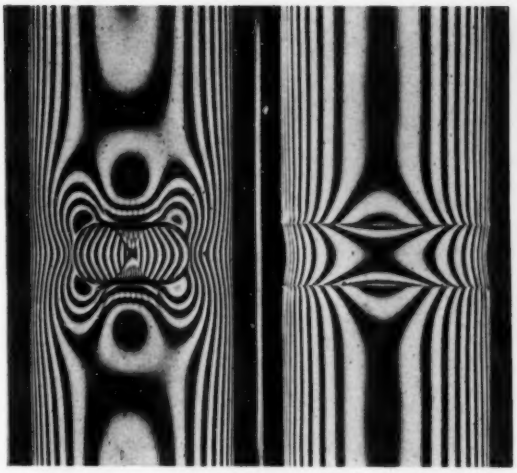


(a)



(b)

FIG. 2. (a) Frozen stress pattern of a circular shaft with transverse hole. $D=1.069''$; $r=0.132''$; $r/d=0.164$ (initial dimensions); $P=22.15$ lb.; f (nominal) $=1.42$ p.s.i. shear; $N=9.4$ fringes. (b) Stress pattern showing the absence of large stresses in the plane of a transverse section of shaft, $t=0.257''$.



(a) (b)

FIG. 3. (a) Frozen stress pattern of a circular shaft with a transverse hole in tension. Direction of ray inclined with respect to axis of hole 20° approximately. $D=0.963''$; $r=0.131''$; $r/d=0.187$; $P=20.95$ lb.; f (nominal) $=1.60$ p.s.i. shear; $N=9.0$ fringes. (b) Frozen stress pattern of the same shaft with the direction of the light ray perpendicular to the axis of the hole.

Methods and Models

The method employed was that of freezing or annealing the stresses into a Bakelite model and subsequently cutting it into thin slices for individual observation in the polariscope.[2] The model shafts were all of Bakelite BT-61-893, approximately one inch in diameter and seven inches long. They were heated to a temperature of about 260°–265°F for two to three hours, and slowly cooled under load to room temperature, the time of cooling being about 12 to 14 hours. Typical frozen stress patterns of the shafts are shown in Figs. 1–3. The slicing was performed on a small milling machine with either circular saws or end mills. Two circular saws were used: one of diameter 4 inches and $\frac{1}{16}$-inch thickness, the other $2\frac{1}{2}$ inches in diameter and $\frac{1}{32}$ inch in thickness. All sections were made perpendicular

[2] For a discussion of the theory of the frozen stress pattern see Max M. Frocht, *Photoelasticity*, Vol. I, 1.07 through 10.11 (John Wiley & Sons, New York); or M. Hetenyi, "The fundamentals of three-dimensional photoelasticity," J. App. Mech. (December, 1938).

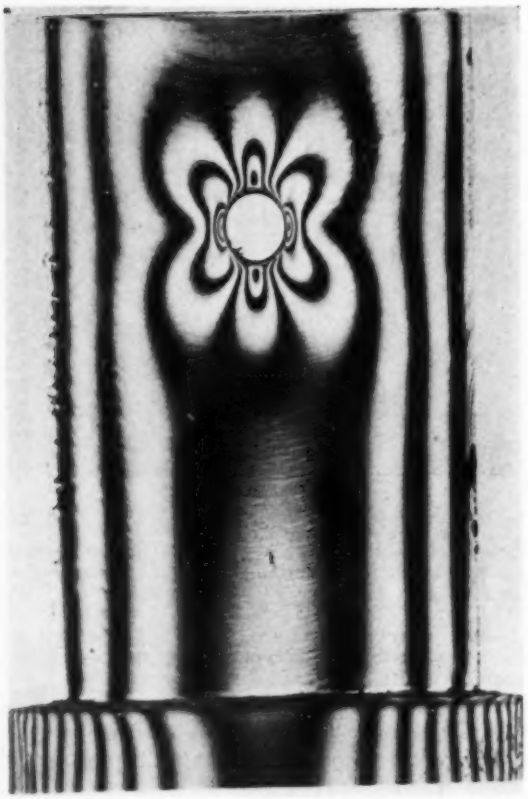


FIG. 4. Frozen stress pattern of a section of a circular shaft showing effect of time on boundary stresses. $2r=0.124''$; $D'=0.865''$; $t=0.234''$; $r/d=0.067$; n_{max} (extrapolated) 7.8 fringes.

to the axis of the hole, so that each section represented a thin plate with a central transverse circular hole. The general dimensions of the models and the manner of slicing will be shown later.

Boundary Difficulties

The greatest difficulty in studying stress concentrations photoelastically is the difficulty of making the boundaries of the model and the boundary stresses visible, especially at the sources of stress concentration. These difficulties were successfully disposed of in two-dimensional problems.[3] In three-dimensional work, however, they recur again and we are once more confronted with boundary disturbances and the necessity for extrapolation.[4] The accuracy of three-dimensional photoelastic work is, therefore, inherently lower than that of corresponding two-dimensional investigations.

The causes of the present boundary disturbances are two: time stresses resulting from the relatively long period of time needed for heating, cooling, and machining, and the inevitable deformations and distortions which are present in three-dimensional problems. The stress pattern of Figs. 4 and 6 show both types of boundary disturbances. The time stresses manifest themselves in changing the shapes of fringes at the boundary and therefore also the fringe orders, Fig. 4. The geometric distortions or deformations of the hole are responsible for the large shadows or space effect around the hole, Fig. 6. These shadows result from the fact that the circular hole not only becomes elliptical but that these ellipses have different dimensions at different sections across the hole. The shape of the hole after deformation is shown in Fig. 5, which gives transverse and longitudinal sections through the hole. It will be observed that in the transverse section the boundaries of the hole are concave, and that in the longitudinal section they are convex.

The use of thin plates or sections effectively removes the shadows or space effects. We have used slices as thin as $\frac{1}{16}$ of an inch with good results. When very thin plates are employed, it may become necessary to resort to the doubling polariscope, or the Babinet compensator.

Method of Loading

Considerable difficulty was encountered in applying axial loads. A variety of mechanical clevices of the universal type were tried but all left traces of bending or twisting, or both. The best alignment was obtained when the metal clevices were replaced by loops of twine or strong string. These were formed into closed bands, put through the holes in the bars of the loading frame, and the ends slipped over the pins passing through the bar. It should perhaps be added that

[3] For an explanation of the causes of boundary vagueness and the proper technique of making them clearly visible, see Max M. Frocht, *Photoelasticity*, Vol. I, 11.4 through 11.8, reference 2.

[4] See paper by E. E. Weibel, "Studies in photoelastic stress concentrations"; also A. M. Wahl and R. Beeuwkes, "Stress concentrations produced by holes and notches" Trans. A.S.M.E. **56** (August, 1934).

even greater difficulties are encountered in producing uniform compression. The effect of bending can be further eliminated by placing the pins which carry the loads at right angles to the open transverse hole under investigation. Such an arrangement of the pins combined with flexible strings gave excellent results. This is clearly shown in Fig. 1, which gives an elevation and side view of the whole shaft taken through Halowax oil. It will be observed that the fringe order along the axis is the same in both views of Fig. 1. The transverse section of Fig. 2(b) shows the absence of normal stresses in the cross section of the shaft.

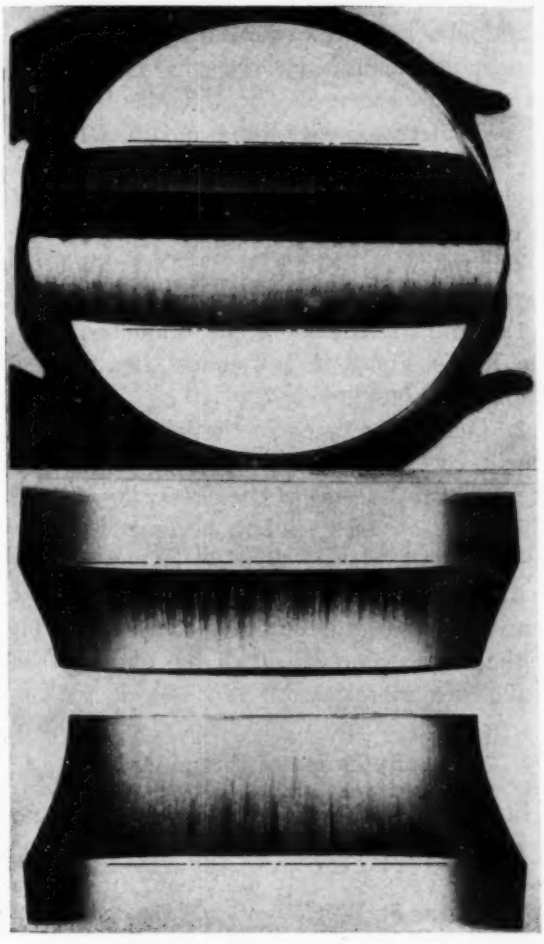
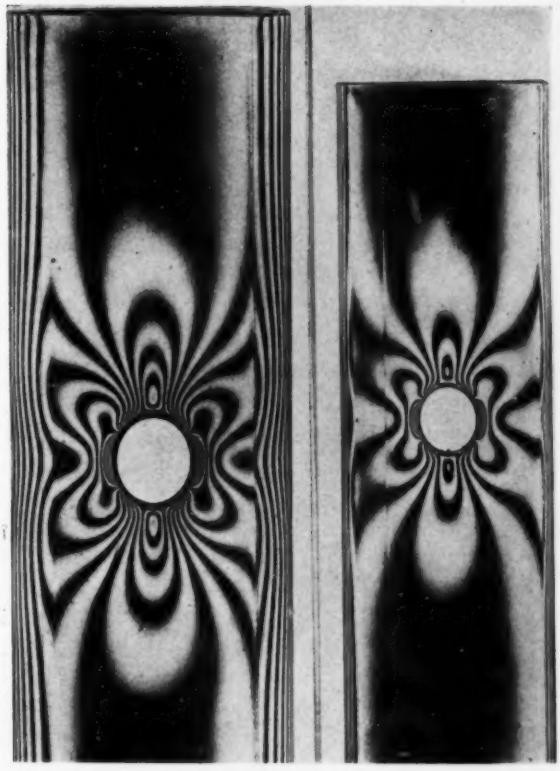


FIG. 5. Photographs showing shape of hole after deformation. In the transverse section the inside diameter is greater than at the ends; in the longitudinal section the converse is true. Obtained from a Plexiglas shaft, diameter ¾ inch approximately.



(a) (b)

FIG. 6. Frozen stress pattern of a section of a circular shaft with a transverse hole obtained by milling off equal amounts from opposite sides illustrating the "difference method." $D=0.963''$; $r=0.131''$; $r/d=0.187$.
(a) $t=0.635''$; $n_{max}=17.8$ fringes.
(b) $t=0.401''$; $n_{max}=11.9$ fringes.

Method of Slicing

Two methods of slicing were used, which we refer to as the *plate method* and the *difference method*. In the plate method thin plates are actually sawed out of the model and each plate is examined in the polariscope. In the difference method the shaft is not sawed into plates but is gradually reduced in thickness, and the difference between the stresses or fringe orders at a point between successive patterns is taken to represent the stress or fringe order in a plate of a thickness equal to that of the materials removed, Figs. 6 and 7.

The difference method is in a sense equivalent to making sections with a knife of zero thickness. Were it not for the inevitable boundary vagueness resulting from the three-dimensional deformations which increases with the thickness, the dif-

FIG. 7. Frozen stress pattern of a thin plate cut from a circular shaft. $D=0.963''$; $r=0.131''$; $r/d=0.187$; $t=0.1875''$; $n_{max}=5.7$ fringes.

ference method would have much in its favor. No material, and therefore no pertinent part of the model, would be lost in slicing. The stress patterns could be made to have a rather large number of fringes at relatively small stresses and so the changes in the curvature at the sources of stress concentrations could be materially reduced. However, all these advantages seem to be offset by the accumulation of space effects or shadows and the subsequent necessity for large extrapolations. The disadvantage of the plate method lies in the small number of fringes that can be frozen into it without large deformations. The boundaries in thin plates are, however, relatively sharp and at the present moment, we find it preferable to work with such plates, even though the number of fringes may be limited. The plate method is also somewhat shorter and results in smaller time stresses during the machining. The whole shaft can be sliced without being removed from the vise in the machine and the plates rapidly photographed. The results which we are presenting below were obtained from a combination of both methods.

Time Stresses

It is known that Bakelite models free from external loads will in a period of several hours develop internal stresses. It has been shown that the formation of these stresses is connected with the absorption of moisture from the air.[5] These stresses, which are known as time stresses, are compressive on the outer surface and tensile in the interior. It has been found that the depth of penetration of these compressive stresses is a function of several variables, such as time and atmospheric conditions. Depths as high as $\frac{1}{8}''$ have been observed. The magnitude of these stresses increases with time and in a few days may reach an intensity of 4 to 6 fringes per inch of thickness. In a period of several months these intensities may be doubled.[6]

The following are additional facts about time stresses, pertinent to our discussion.

1. Time stresses along one diameter, such as the Y axis, Fig. 8, do not cancel, i.e., are not in equilibrium.[7]

2. The total tension in the interval $2a$ of the tensile core is greater than the compression in the interval $2(R-a)$ of the outer ring, Fig. 8.

3. Time stresses tend to become released or redistributed when a section is cut from the shaft. In particular, if a transverse disk is cut from the shaft and the outer compressive layer of the skin is removed, then the tensile stresses in the core are released.

Determination of Fringe Orders

In two-dimensional photoelasticity we use one of two basic methods for the determination of

[5] See papers by Walter Leaf on "Time edge effect," Fifteenth Eastern Photoelasticity Conference, held at Massachusetts Institute of Technology (June, 1942); also "Additional data on time edge effect" presented at the Sixteenth Conference.

[6] See Figs. 11.1 and 11.13 in Max M. Frocht's *Photoelasticity*, Vol. I.

[7] A proof for this statement is given in the Appendix.

fringe orders. One method may be called static, the other dynamic. There are stress patterns which contain singular or isotropic points, i.e., fringes of order zero. A beam in pure bending is a typical illustration. In such cases the fringe order at any point can be determined from a single or static photograph, provided, of course, we also know the manner of pattern formation. Generally, however, the fringe of zero order is not retained in a stress pattern and the formation may be rather complicated. In such cases the most practical method to determine the fringe order at a point is by repeated loading and unloading and continuous observation of fringe formation. We may call this the dynamic method. This method can obviously not be employed in a frozen stress pattern. The static method will work in some, but not in other cases. However, it is always possible to determine the fringe order at any point with great accuracy by compensation.

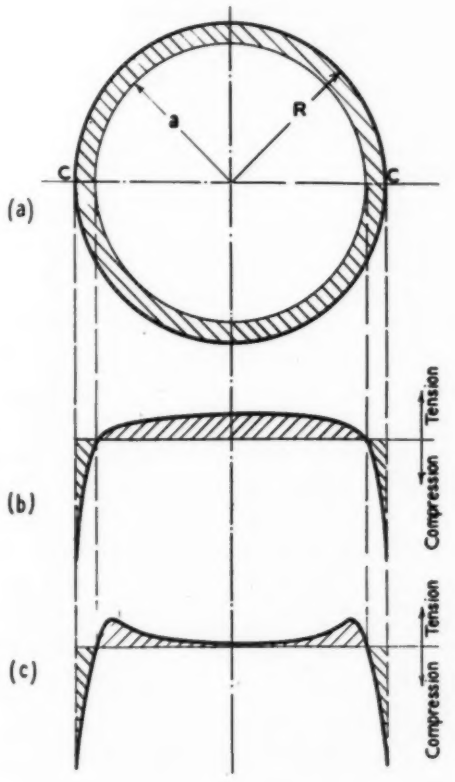


FIG. 8. Sketch showing probable distribution fo time stresses.

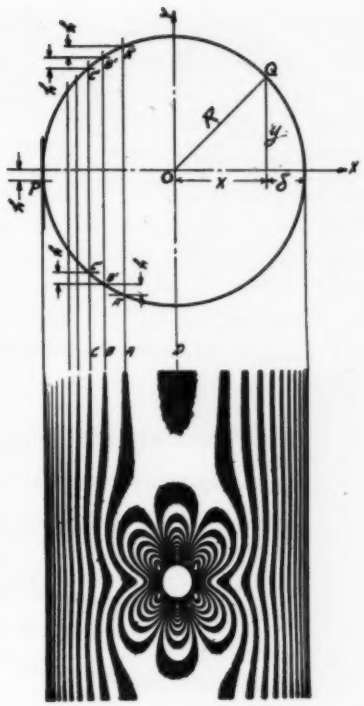


FIG. 9. Sketch showing geometric method for the determination of fringe order.

The Compensator Method

This method is basically the same as that used by Professor Coker for the evaluation of $(p-q)$. Using white light as a source we superimpose a uniform tension or compression over the point under inspection until a dark zone forms. From the order of the isochromatic or magnitude of the stress in the compensator strip at compensation the fringe order at that stress can be readily found. Using white light the load on the compensator strip is measured at compensation. The white light is then replaced by a monochromatic source and the fringe order in the compensating strip at that load determined. The fringe order in the compensating strip at compensation equals the fringe order of the compensated point. The usual compensators are made for a small range of stress and are not suitable for this purpose. However, the compensation can easily be affected by inserting a small tension specimen of arbitrary thickness into the main straining or loading machine, placing the annealed or frozen shaft (or other body) next to it

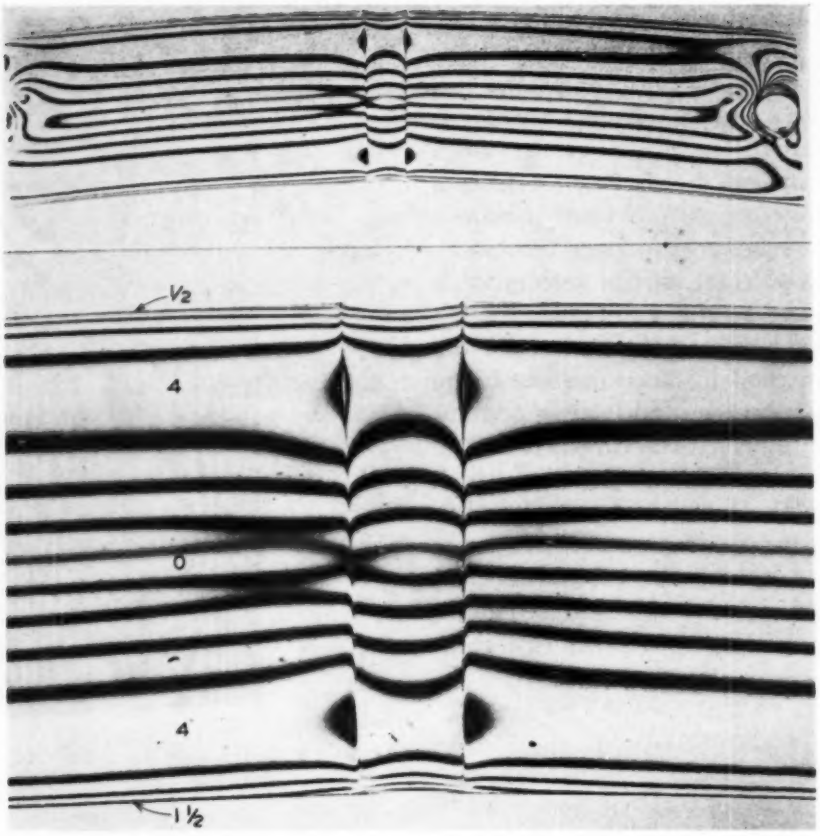


FIG. 10. Frozen stress pattern of a circular shaft with an open hole in pure bending. $D=0.980''$; $r=0.110''$; $r/d=0.145$; $N=4$ fringes. Note spreading of fringes on tension side. (a) Whole shaft; (b) Enlargement of central portion.

at proper orientation, and measuring the stress in the tension specimen at the instant compensation is obtained.

In the case of shafts in tension the fringes at the edges are rather crowded. Because of this it will be found more convenient to compensate the region along the axis of the shaft, where the fringes or isochromatics are broadest.

The Geometric Method

In circular shafts in uniform tension another method, which does not rest on compensation, may be followed. Let A, B, C, Fig. 9, represent the position of three successive fringes of order n_i, n_{i-1}, n_{i-2}. We project these points on the diameter X of the circle representing the cross section of the shaft, and draw chords $A'-A'$, $B'-B'$, $C'-C'$, Fig. 9. The difference $2h$ between the successive lengths of these chords is the thickness needed to produce one fringe. Clearly then, if a line be drawn parallel to the X axis, Fig. 9, distant h from it, this line would locate a point P on the circumference of the circle where fringe one must be placed. It will be noted that this fringe is extremely close to the edge. The order at any other point is obtained by dividing the chord through that point, such as $A'-A'$, by $2h$.

Furthermore, denoting the maximum fringe order at the center 0, Fig. 9, by N, and that at any other point Q by n, and assuming a uniform tension, it is easy to show that the distance δ from point Q to the nearest edge of the shaft is given by

$$\delta=R[1-(N^2-n^2)^{\frac{1}{2}}/N]. \qquad (1)$$

From this equation a theoretical stress pattern could readily be constructed for any assumed

value of N, and compared with the photoelastic pattern. In this way the true value of N could be found. Assuming N to be 10, a simple calculation shows that

$$\delta_1 = 0.00 \quad \text{approximately}$$
$$\delta_2 = 0.02R \quad "$$
$$\delta_3 = 0.046R \quad "$$

in which the subscripts denote the fringe orders. It is seen that the distances from the edge of the shaft to the first three fringes are extremely small, and could readily be lost in the photography or printing. Thus, if one were to rely on this equation he might be led to conclude that in Fig. 1 the first visible fringe near the edge is say 2, instead of 1, making $N = 10.4$ instead of the given 9.4. However, this method of fringe order determination is unreliable, owing to the presence of compressive time stresses in the outer layers of the shaft, the effect of which is to increase the values of δ near the edges, i.e., to spread the fringes apart. This effect is clearly demonstrated by the stress pattern of a shaft in pure bending, Fig. 10.

Here the fringe of zero order forms the neutral surface and appears clearly in the central portion of the beam. This provides a reliable origin from which all other fringe orders can be determined. Inspection of Fig. 10 shows that on the tension side we can clearly identify the fringe of order $\frac{1}{2}$ and even zero, whereas on the compressive side the lowest fringe which is definitely visible is approximately of order $\frac{3}{2}$. The effect of the initial compression on the edges is thus to spread the fringes apart and to make them more easily identifiable on the tension side. Figure 11 provides further and more direct evidence of this effect. The stress patterns shown are of the same section of the shaft but taken four months apart, approximately. Due to the growth of the time stresses there is not only a definite spreading of the fringes at the edges, but the edge stresses have actually changed from tension to compression.

The geometric method would give accurate results if, and only if, a portion of the shaft could be found where the state of stress is absolutely uniform. Such an ideal condition is practically impossible to obtain, even if the applied loads are truly axial and the shaft is free from bending or twisting. The time stresses, which are compressive on the outer surface and tensile in the interior, combine with the uniform tension to produce a non-uniform state of stress, with the result that the fringe order ceases to be proportional to the distance traversed by the light. Because of these unavoidable disturbances in the uniformity of the stress distribution, the geometric method and Eq. (1) may lead to considerable errors in the fringe order. The most reliable method for the determination of fringe orders in the case under consideration is therefore that of compensation. This is the method used in this investigation.

Fringe Value

The presence of the time stresses also complicates the determination of fringe values. Were it not for these stresses the fringe value F for a thickness equal to the diameter of the shaft could be readily found once the fringe order N at the center is accurately determined. Thus, denoting the load and cross-sectional area by P and A, respectively, we have

$$2F = P/AN.$$

Recalling that the fringe value of the material $f = FD$, where D is the diameter of the shaft, we

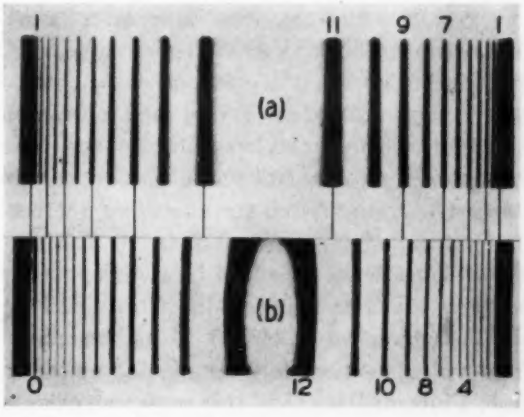
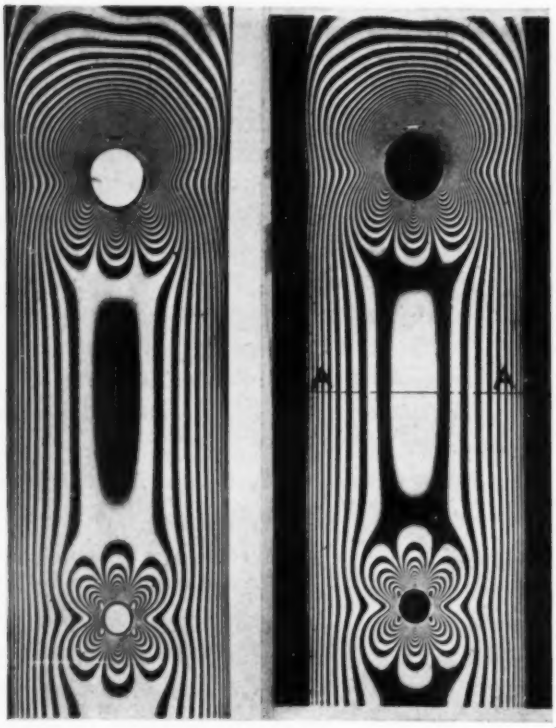


FIG. 11. (a) Enlarged photograph of transverse section A–A, Fig. 12 (b) taken at the time the test was made. $N = 11.7$; (b) Enlargement of same section taken about four months after test, showing spreading of the fringes at the edges of the shaft and an increase in maximum nominal fringe order from 11.7 to 12.25. Note that the time stresses have changed the edges of the shaft from tension to compression.



(a) (b)

FIG. 12. Frozen stress patterns of a circular shaft with a transverse hole, light and dark backgrounds. $D=1.055''$; $r=0.124''$; $r/d=0.067$ (initial dimensions); $P=28.47$ lb.; f (nominal) $=1.49$ p.s.i. shear; $N=11.7$ fringes. See Fig. 11.

get

$$f=PD/2AN. \tag{2}$$

However, the effect of the time stresses is to produce an additional tension over a relatively large central core. In a one-inch shaft this core would be approximately $\frac{15}{16}$ or $\frac{7}{8}$ inch in diameter. As already stated, the resultant of the tensile stresses over a diametral section is greater than that of the compressive stresses. The net result of this state of stress is to increase the fringe order N over what it would be if there were no time stresses. The fringe value f obtained from Eq. (2), which we will refer to as the nominal fringe value, is therefore smaller than its true value. Thus, in Part II of this paper we calculate nominal fringe values and find one as low as 1.42 and another 1.49, which are considerably smaller than the fringe value of 1.59 p.s.i. previously found for two-dimensional flat bars.

However, it can be shown that the true fringe value is still essentially the same as that previously established, i.e., 1.59 p.s.i. In order to find this true fringe value it is necessary to eliminate or determine the time stresses.

The time stresses can be eliminated by removing the compressive layer from a transverse disk cut from the shaft. This can be easily done by turning. The fringe value determined from such a disk cut from a region in which the load stresses are uniform approaches the true value of 1.59 p.s.i. It is this difference between the true and nominal fringe values which makes it impossible to obtain the exact fringe order at a point in the actual stress pattern from the true fringe value.

PART II

Stress Concentrations in Shafts with Transverse Holes in Tension

Case I

We first consider a shaft with a relatively small hole. A stress pattern of the shaft is shown in Fig. 12. The manner of slicing is shown in Fig. 13. The important dimensions are given below:

$D=$diameter of shaft before loading $=1.055''$, $D_f=$diameter of shaft after loading $=1.036''$, $2r=$diameter of hole $=0.124''$, $d=(D-2r)=0.931''$, $r/d=0.067$ for initial dimensions, $r/d=0.1$ for final dimensions approximately, $A=$transverse area of strained shaft after loading $=0.842$ sq. in., $A_{\min}=$net or minimum transverse area through hole $=0.713$ sq. in., $A/A_{\min}=1.185$, $P=$total load $=28.47$ lb., $n'=$fringe intensity at a point on the tension edge of the hole, i.e., the number of fringes per inch of length of lines $A-A$, Fig. 13. This we define by the relation,

$$n'=\lim_{\Delta t\to 0}\frac{\Delta n}{\Delta t}=\frac{dn}{dt},$$

in which Δn is the number of fringes produced by a plate Δt inches thick when the light travels parallel to this thickness, $N=$maximum or total number of fringes in the central part of the solid shaft above the hole where the stress is uniform. This fringe order was determined by compensation. It includes the effect of the time stresses. This number of fringes was found to be 11.7 and

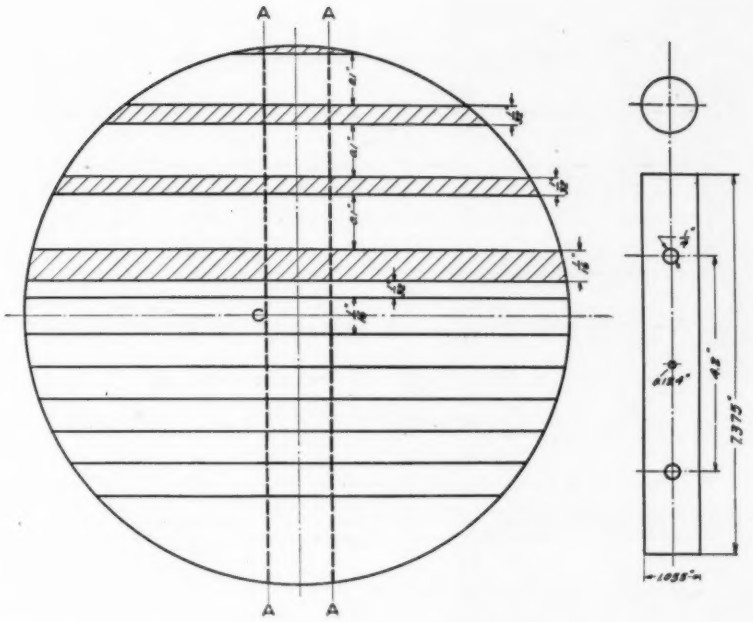


FIG. 13. Drawing showing manner of slicing shaft shown in Fig. 12. Material removed by cutter is shaded.

it occurs along the diameter of the shaft which is parallel to the axis of propagation of the light, f = fringe value of material in p.s.i. shear stress:

$$f=\frac{PD}{2AN}=\frac{28.47\times1.036}{2\times0.842\times11.7}=1.49 \text{ p.s.i.},$$

n'_{unf} = fringe intensity in the region of pure tension, i.e., the number of fringes per inch:

$$n'_{\text{unf}}=\frac{N}{D}=\frac{11.7}{1.036}=11.3 \text{ fringes per inch},$$

n'_{hn} = the average intensity across the minimum or net transverse area through the hole. Since

$$n'_{\text{unf}}=P/2Af,$$

and

$$n'_{hn}=P/2A_{\min}f, \quad n'_{hn}=n'_{\text{unf}}A/A_{\min}, \qquad (3)$$

$$=11.3\times1.185=13.4 \text{ fringes per inch.}$$

The last fringe intensity may also be found from the average stress in psi divided by the known fringe value. Thus

$$n'_{hn}=\frac{P}{2A_{\min}f}=\frac{28.47}{2\times0.714\times1.49}=13.4 \text{ fringes per inch.}$$

Curve I of Fig. 14 shows the experimental values of n' for the tension edge of the hole. These intensities were obtained by examining separate plates or from the differences between plates. The significant aspect about these fringe intensities, from the point of view of the engineer, lies in the fact that they increase in magnitude as we approach the boundary of the shaft. The stress patterns of Fig. 15 which were obtained from plates of equal thickness (0.1″) clearly show this increase in the fringe intensity. From this we conclude that the maximum tensile stresses around the hole lie on the outside surface of the shaft. In the case under consideration the increases in the fringe intensities between the center of the shaft and the outside boundary is about 17 percent, the corresponding values of n' being 32 and 37.5, approximately.

Curve II of Fig. 14 shows the total number of fringes for a plate as a function of the thickness t measured from the center of the shaft. This curve of n *vs.* t is obtained from curve I by taking the area under it. For, by definition

$$n'=dn/dt,$$

and therefore

$$n=\int_0^t n'dt,$$

i.e., n is the area under the n' curve.

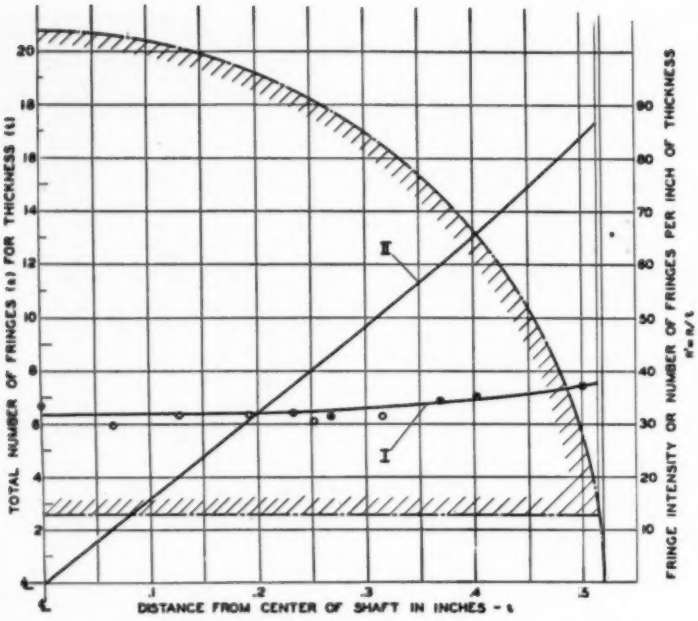

TOTAL NUMBER OF FRINGES (n) FOR THICKNESS (t)
FRINGE INTENSITY OR NUMBER OF FRINGES PER INCH OF THICKNESS
n'=n/t
II
I
DISTANCE FROM CENTER OF SHAFT IN INCHES - t

FIG. 14. Curves showing the variation in tensile stress intensity along edges of hole in shaft of Fig. 12. Curve I gives fringe intensity n', i.e., the number of fringes per inch of thickness. Curve II gives total number of fringes n as a function of thickness t.

Factor of Stress Concentration k_3

We define the factor of stress concentration for the shaft in the usual manner, i.e.,

$$k_3=\sigma_{max}/\sigma_{nominal},$$
$$=\sigma_{max}/(P/A_{min}),$$

so that

$$\sigma_{max}=k_3(P/A_{min}).$$

In the case under consideration

$$n'_{max}=37.5 \text{ fringes per inch,}$$

and in terms of fringes

$$\text{nominal}=n'_{hy}=13.4 \text{ fringes per inch.}$$

Hence

$$k_3=37.5/13.4=2.80.$$

It should be noted that the curvature of the hole is somewhat altered by the freezing process, and therefore the ratio r/d changes. For the shaft under consideration this ratio changed from $r/d=0.067$ to $r/d=0.1$, approximately. The factor k_3 found above must be interpreted as referring to the modified value of r/d. It may be added that this is only a nominal correction based on the deformations produced at the hole in the central transverse section of the shaft.

Case II

$D=1.069''$, $D_f=1.054''$, $2r=0.265''$, $d=(D-2r)=0.804''$, $r/d=0.164$ initial dimensions, $r/d=0.207$ corrected dimensions, $A=0.874$ sq. in., $A_{min}=0.599$ sq. in., $A/A_{min}=1.46$, $P=22.15$ lb., $f=1.42$ p.s.i. shear stress (3 hours at high temperature), $N=9.4$,

$$f=\frac{PD}{2AN}=\frac{22.15\times1.054}{0.874\times2\times9.4}=1.42,$$

$$n'_{unf}=N/D_f=8.91 \text{ fringes per inch, } n'_{hy}=13.$$

The second case not only corroborates the first but the increase in the stress intensity is even more pronounced. Stress patterns of the shaft as a whole appear in Figs. 1 and 2 and clearly show a state of symmetrical tension. This shaft was cut into five plates of equal thickness of $0.177''$ each, Fig. 16. The stress patterns of these are given in Fig. 17. They clearly demonstrate the rise in stress intensity toward the edges of the hole. Inspection of the curve in Fig. 16 shows the symmetrical nature of this rising intensity. The factor of stress concentration k_3

for this shaft is

$$k_3 = n'_{max}/n'_{Nu},$$
$$= 33/13 = 2.54.$$

Attention is called to the relatively small time stresses visible around the hole. For example, the photographs clearly show the presence of the four singular points on the circumference of the hole. It is probable that this is in a measure due to a release of the time stresses by the cutting process itself.

Case III

Another shaft with a large hole was next studied. The dimensions and the loads are given below, and a stress pattern of a portion of the shaft containing the hole is shown in Fig. 18.

$D = 1.116''$, $D_f = 1.108''$, $2r = 0.438''$, $d = (D - 2r) = 0.678''$, $r/d = 0.323$ for initial dimensions, $r/d = 0.385$ for corrected dimensions, $A = 0.964$ sq. in., $A_{min} = 0.491$ sq. in., $A/A_{min} = 1.96$, $P = 18.48$ lb., $N = 7.1$,

$$f = \frac{PD_f}{2AN} = \frac{18.48 \times 1.108}{0.964 \times 2 \times 7.1} = 1.49 \text{ p.s.i. shear stress,}$$

$n'_{unf} = N/D_f = 6.41$ fringes per inch, $n'_{Nu} = n'_{unf}\dfrac{A}{A_{min}}$ $= 12.56$ fringes per inch.

The manner of slicing is given in Fig. 19, and a typical stress pattern of a plate is shown in Fig. 20. As in the preceding cases the fringe intensity increases as we approach the boundary. Curves of n and n' are given in Fig. 21.

The factor of stress concentration for this case is

$$k_3 = n'_{max}/n'_{Nu},$$
$$= 31.25/(6.41 \times 1.96) = 2.40.$$

The experimental results of all three cases are summarized in Fig. 22. Curve I gives the final values of k_3 corrected approximately for the change in curvature. Curve III gives the factors of stress concentration for the stress intensity at the center of the shaft. For example, in Fig. 16 this stress intensity is 26.5 at the origin 0. These values should not be used in design. Curve II is the well-established curve of k_2 for the corresponding two-dimensional plates having the same ratios r/d as the shafts.

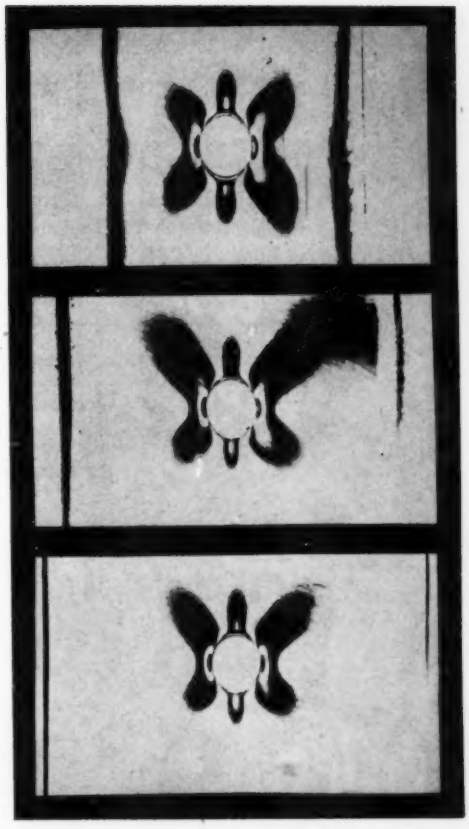


FIG. 15. Frozen stress patterns of thin plates cut from shaft down in Fig. 12 showing increase in the tensile stress intensity, as the distance from the center of the shaft increases. Each plate is 0.1" thick.

Top: n_{max} (extrapolated) 3.6; $D'(max) = 0.650''$; $r = 0.062''$; $r/d = 0.067$.
Center: n_{max} (extrapolated) 3.2; $D'(Av) = 0.807''$; $r = 0.062''$; $r/d = 0.067$.
Bottom: n_{max} (extrapolated) 3.1; $D'(Av) = 0.965''$; $r = 0.062''$; $r/d = 0.067$.

PART III

Interpretation of Results. Independent Laminar Action

Inspection of curve I, Fig. 14, shows that in the interior, or central part, of the shaft the fringe intensity is constant. This is clearly true up to a point where $t = 0.25''$ measured from the center of the shaft. Even beyond that point the rise is rather gradual. This shows that the interior part of the shaft behaves as if it consisted of vertical laminae or plates, transverse to the

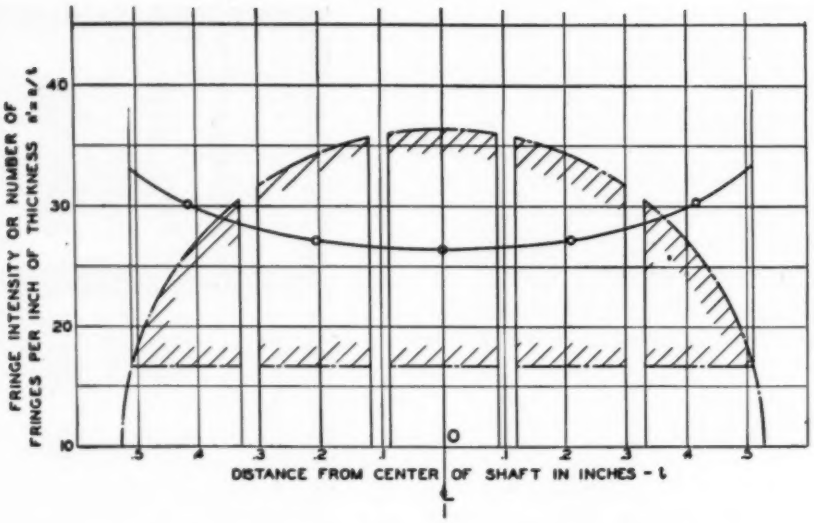


FIG. 16. Curve showing the variation of the tensile stress intensity along the edge of the hole. Figure also shows the manner in which the plates were cut from the shaft. $D=1.069''$; $r=0.132''$; $r/d=0.164$; $t=0.177$ for each plate; $n'_{\min}=26.5$; $n'_{\max}=33.2$. See Fig. 3 for stress pattern of whole shaft.

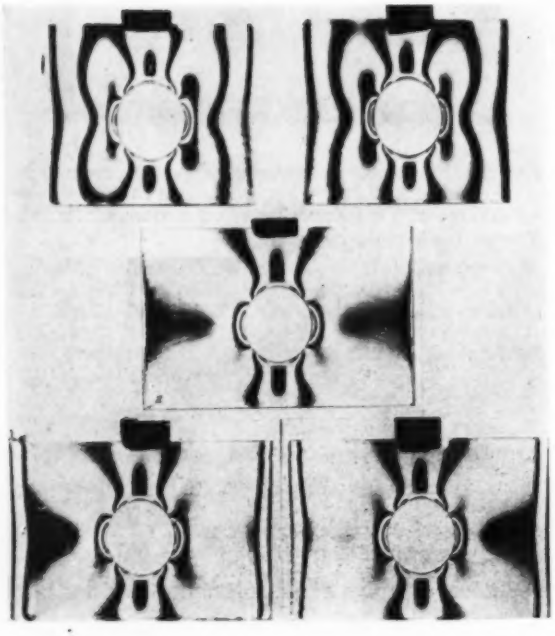


FIG. 17. Frozen stress patterns of plates cut from a solid shaft in Fig. 16 showing the variation in the stresses along the hole as a function of the distance from the center of the shaft. The upper row shows the two end plates of the shaft. The center pattern is that of the center plate, and the lower row gives the two intermediate plates. It will be observed that the fringe order reaches a maximum of nearly five in the two end plates, while it is only four in the center plate. The thickness for each is about 0.177". Attention is called to the remarkable clarity of the four singular points around the hole. These, and the pattern as a whole, indicate the absence of marked time stresses.

hole, each acting independently of the other. This is equivalent to saying that each transverse plate acts like a thin plate with a central circular hole in two-dimensional tension and that the factor of stress concentration for each plate, which we designate by k_2, is determined by the ratio r/d', in which d' is the chord or effective width of the plate, Fig. 23.

Assuming such a laminar action the fringe intensities n' along the tensile edge of the hole can be expressed by the equation

$$n' = k_2 n'_{\text{unf}} D'/d' \qquad (4)$$

in which D' and d' are as shown in Fig. 23, and n'_{unf} is the fringe intensity or stress in the region of pure tension. The values of n' given by (4) we will refer to as theoretical intensities.

Figure 24 shows a comparison of the experimental and the theoretical values of n', from which it can be seen that most of the shaft behaves as if it consisted of separate laminae. However, this is more nearly true for small holes than for large ones, i.e., when the holes are large in relation to the diameter of the shaft a smaller part of it acts that way.

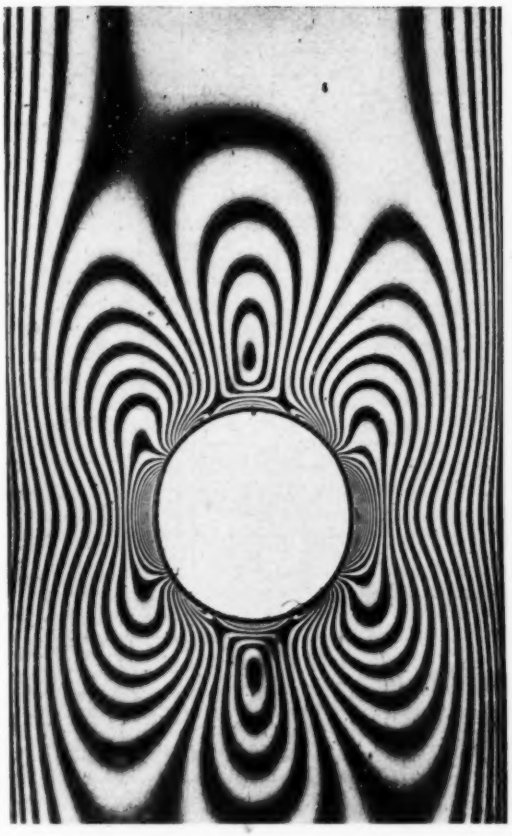


FIG. 18. Frozen stress pattern of a circular shaft with a transverse hole. $D=1.116''$; $r=0.219''$; $r/d=0.323$; $P=18.48$ lb.; f (nominal) $=1.49$ p.s.i. shear.

Relation Between Two- and Three-Dimensional Factors of Stress Concentration

By definition

$$k_3=\sigma_{max}/\sigma_N=n'_{max}/n'_N.$$

Let

$$n'_{max}=(n')_{t=0}\times f(r/d),$$

in which $(n')_{t=0}$ denotes the fringe intensity at the center of the shaft, where $t=0$, and $f(r/d)$ is some function of r/d defined by curves I of Figs. 14, 16, and 21. Now, experiments show that for all practical purposes

$$(n')_{t=0}=k_2 n'_{unf}D/(D-2r). \tag{5}$$

Hence

$$n'_{max}=k_2 n'_{unf}Df(r/d)/(D-2r).$$

Further

$$n'_N=n'_{unf}A/A_{min}.$$

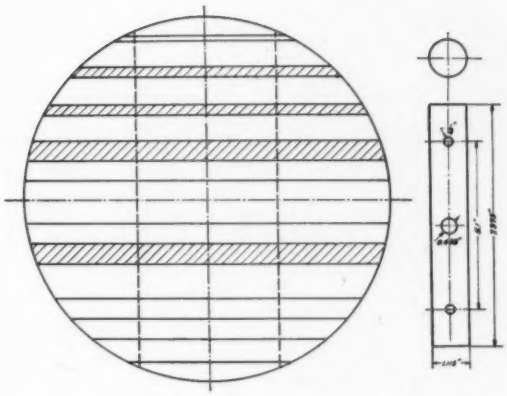


FIG. 19. Drawing showing manner of slicing shaft shown in Fig. 18. Material removed by cutter is shaded.

Hence

$$\begin{aligned} k_3&=k_2Df(r/d)A_{min}/(D-2r)A,\\ &=f(r/d)k_2\times D/d\times A_{mn}/A. \end{aligned} \tag{6}$$

It will be found that for the cases presented in

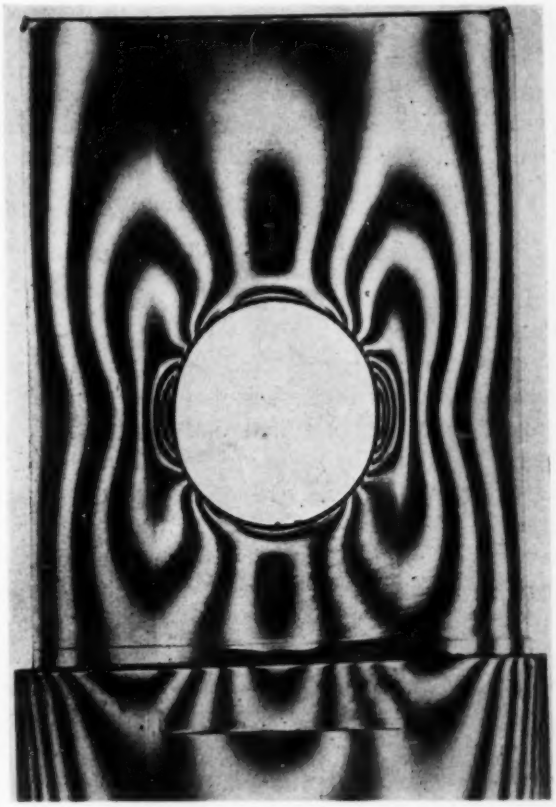


FIG. 20. Frozen stress pattern of a plate cut from the circular shaft with a transverse hole shown in Fig. 18. $D'=1.038$; $r/d=0.323$; $t=0.300''$; $n_{max}=8.1$; f (nominal) $=1.49$ p.s.i. shear.

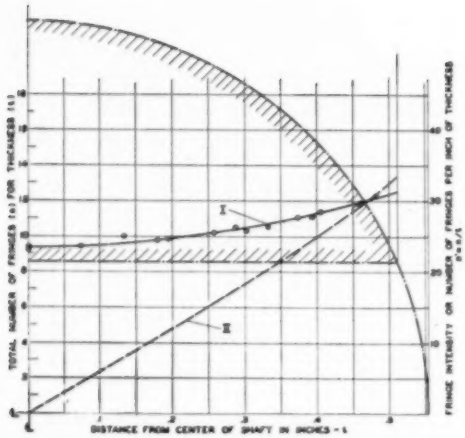

TOTAL NUMBER OF FRINGES (n) FOR THICKNESS (t)
FRINGE INTENSITY OR NUMBER OF FRINGES PER INCH OF THICKNESS
DISTANCE FROM CENTER OF SHAFT IN INCHES - r
I
II

FIG. 21. Curves giving the variation of the tensile stress along edge of hole, and the total number of fringes as a function of the thickness, for shaft shown in Fig. 18. Curve I gives n', curve II gives n.

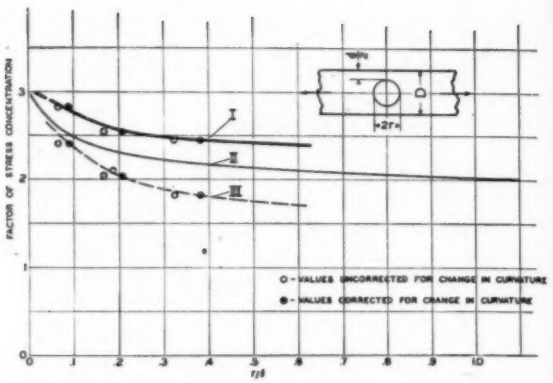


FIG. 22. Summary of experimental results. Curve I gives experimental values of the factor of stress concentration k_3. Curve II gives the two-dimensional factor k_2 for the same r/d. Curve III gives k_3 bases upon the tensile stress of the hole in the center of the shaft. This last value is shown only for comparison and should not be used in design.

this paper the formula

$$k_3=(1.1+7.1r/d)k_2\times D/d\times A_{\min}/A, \quad (7)$$

gives the experimentally determined values of k_3. Hence, for shafts with holes in which r/d does not exceed 0.5, and such cases are probably of greatest practical importance, the above equation will give good approximations for the factors of stress concentration.

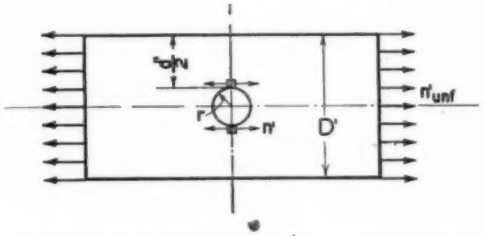


FIG. 23. Drawing showing notation used for shafts with transverse hole.

Discussion of Results and Conclusion

Inspection of curve I, Fig. 22, shows that in the range indicated the values of k_3 do not differ by more than 14 percent approximately from the corresponding two-dimensional factors k_2. It should, however, be noted that these values of k_3 are based on the values of N, uncorrected for the effect of the time stresses. These values are high. If the true values of N were used the average stress in the transverse section through the hole would become smaller. Assuming that the maximum stresses around the hole as given in Figs. 14, 16, and 21 are essentially correct this would mean that our values of k_3 given by curve I, Fig. 22, are low in the ratio of 1.59/1.45, approximately. Our results therefore represent a lower limit for the values of k_3, the probability existing that the true values may be somewhat higher.

The results given by curve I, Fig. 22, may be considered quite reliable in the range of r/d between 0.08 and 0.5, approximately. There is some doubt about the limiting or maximum value of k_3 as r/d approaches zero. Inspection of the curves giving n' in the three different cases, Fig. 25, shows that as the hole gets smaller the curves of n' flatten out and $n'_{\max}$ approaches the value of n' at the center of the shaft. Thus the ratio of $n'_{\max}/n'_{\min}$ is 1.353 and 1.186 for $r/d=0.323$ and 0.067, respectively. It is therefore reasonable to conclude that for a very small hole the limiting value of k_3 is the same as for k_2 which has been established as 3. However, this conclusion is tentative and calls for further investigation.

In further work on these problems efforts will have to be made to eliminate or to reduce the effects of time stresses. The elimination of this annoying phenomenon will probably have to wait until a new material is found which does not possess this characteristic.

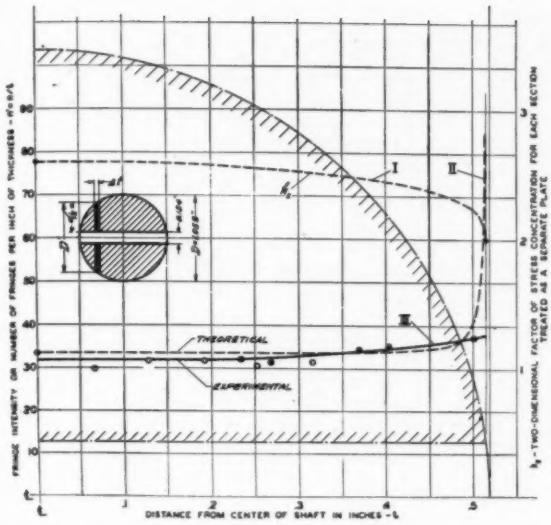


FIG. 24. Curves showing experimental and theoretical values of fringe, or stress intensity n'. Curve I represents two-dimensional factors k_2. Curve II is obtained from the expression $n' = k_2 n'_{unit} D'/(D'-2r)$.

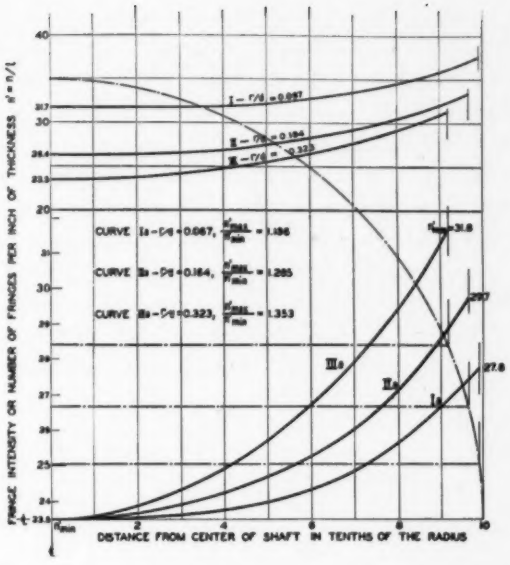


FIG. 25. Curves showing relative rise in fringe intensities (n') for the three cases presented. Curves I, II, and III are taken from Figs. 14, 16, and 21. In order to bring the curves to one origin, the loads on the three shafts were adjusted to produce the same stress at the center of the shaft, i.e., the same n'_{min}. The stress chosen was that of case III in which $n'_{min} = 23.5$. Curves Ia, IIa, and IIIa are the corresponding curves of n'. The percent increase in the stress intensity in each of the three cases is shown in the figure.

APPENDIX

Relation Between Time Stresses

We assume that the time stresses follow some such distribution as shown in Fig. 26. In these distributions there is an outer circular ring in compression, of a depth varying with time, and a much larger central core in tension.

Conditions of equilibrium require that the resultant compression on the ring be equal to the total tension on the core. Denoting the compressive stresses by σ_c and the tensile stresses by σ_t we have

$$\int_a^R \sigma_c 2\pi r dr = \int_0^a \sigma_t 2\pi r dr \tag{8}$$

or

$$\int_a^R \sigma_c r dr = \int_0^a \sigma_t r dr. \tag{9}$$

This equation shows that in a diametral section of the time stresses, Fig. 26, the moments of the compressive and tensile forces about the origin O must be equal.

We observe that in order to have equilibrium among the time forces in a diametral section, Fig. 26, it is necessary that

$$\int_a^R \sigma_c dr = \int_0^a \sigma_t dr. \tag{10}$$

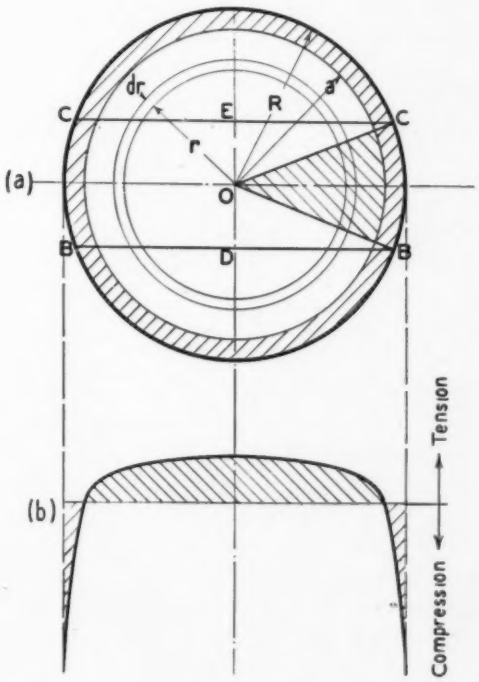


FIG. 26. Sketch showing probable distribution of time stresses across a diametral section. Also relation between tensile and compressive forces.

In view of Eq. (9) the last condition is possible only if r is a constant, which is clearly not the case. It follows that the time stresses along any diametral section such as shown in Fig. 26 are not in equilibrium.

Another way of arriving at the same conclusion is to consider the forces acting on the section *CCBB*, Fig. 26. Clearly the time stresses acting on the curved triangle *OBC* are in equilibrium. It follows that the forces on the triangles *OBD* and *OCE* do not balance. Hence the time stresses on a transverse section of an axial longitudinal plate are not in equilibrium, and upon cutting such a plate from the shaft there must be some release or readjustment among the time stresses.

ACKNOWLEDGMENTS

The author expresses his appreciation to the administration of the Carnegie Institute of Technology for the support of the work. Special acknowledgment is made to Mr. John R. McDowell, mechanician in the Photoelastic Laboratory, for his assistance and cooperation.